DOCUMENT RESUME

ED 219 717 CS 006 716

AUTHOR Barnes, Nancy Marie Van Stavern; Barnes, James Albert, Jr.
TITLE Vocabulary Studies: Summarized, Reviewed, Critiqued, and Offered in an Historical Perspective.
PUB DATE [75]
NOTE 101p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Child Language; Literature Reviews; Oral Language; Primary Education; *Reading Research; *Structural Analysis (Linguistics); Textbook Content; *Vocabulary; *Vocabulary Development; *Word Frequency; *Word Lists

ABSTRACT

To offer a historical perspective on vocabulary studies, this paper presents summaries, reviews, and critiques of vocabulary studies in chronological order of publication. The only variance from this format are published rebuttals and/or critiques, which are presented immediately following the original studies. The paper is divided into two parts. Part one presents the "Early Studies: Pre-1950" and contains the studies that were published before 1950. Part two presents "More Recent Studies: Post-1950" and contains the studies published since 1950. No attempt is made to classify the studies except in historical sequence. Basically the studies are concerned with the reading, writing, and spoken vocabulary of young children, in grades one, two, and three. Also, "original" and primary sources are used, not secondary. The paper concludes with a summation of the studies, noting that while the earlier studies were mainly concerned with an analysis of vocabulary through frequency word counts and comparison of previously published lists, the last three decades from the 1950s to the present give evidence of a shift in emphasis to a linguistic analysis of the language. An extensive bibliography is included. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219717

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

VOCABULARY STUDIES:

SUMMARIZED,

REVIEWED,

CRITIQUED,

AND OFFERED IN AN HISTORICAL PERSPECTIVE

Nancy Marie Van Stavern Barnes
Thermalito Elementary School District
1050 Sierra Ave.
Oroville, CA 95965

and

James Albert Barnes, Jr.
Palermo Elementary School District
2261 Esperanza Ave.
Palermo, CA 95968

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Nancy Marie Van Stavern Barnes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS506716

Nancy M. Barnes
James A. Barnes, Jr.
5414 Pentz Rd Rt. 4
Paradise, CA 95969

TABLE OF CONTENTS

	Page
Introduction	1
EARLY STUDIES: PRE-1950	2
The 1900's	
Housh, E.T. (1918)	2
The 1920's	
Packer, J. L. (1921)	3
Thorndike, Edward L. (1921,a.b.)	3
critiques of Thorndike:	
Gates, Arthur I. (1926, a.)	5
Dolch, Edward W. (1928; 1949)	6
Dale, Edgar (1931; 1941)	6
Selke, Erich, and G.A. Selke (1922)	7
Gray, C. T. (1925)	7
Kircher, W. Gray, (1925)	8
Gates, Arthur I. (1926, a.)	8
Gates, Arthur I. (1926, b.)	10
critique of Gates:	
Wheeler and Howell (1930)	11
Horn, Ernest (1926)	11
Horn, Madeline (1927; 1928)	12
Dolch, Edward W. (1927)	13
Dolch, Edward W. (1928, b.)	14
Dolch, Edward W. (1928, a.)	16
Fennell, Ethel (1928)	16
critique of Fennell:	
Hockett, J. A. (1937)	17
Sister Irmina (1929)	17
Selke, Erich (1929)	18
The 1930's	
Wheeler, M. E. and Emma Howell (1930)	18
Dolch, Edward W. (1930)	19
Selke, Erich (1930)	20
Dale, Edgar (1931)	20
Harring, Sidney (1931)	21
Fitzgerald, James (1934)	21
Thorndike, Edward L. (1931)	22
Gates, Arthur I. (1935)	23
Dolch, Edward W. (1936, a.)	23
Dolch, Edward W. (1936, b.; 1939; 1948)	24
critique of Dolch:	
Stone, Clarence R. (1950)	28
Hockett, John A. (1936)	28
Buckingham, B. R. and E. W. Dolch (1936)	29

Continued: The 1930's

Page

Gray, C. T. (1936) 30
McKee, P. (1937) 30
Hockett, J. A. (1937). 31
Hockett, J. A. (1938). 32
Fitzgerald, James (1938) 33
Betts, I. A. (1939). 33
Fielstra, C. and F. D. Curtis (1939) 35

The 1940's

Seegers, J. C. (1940). 36
critique of Seegers:
Thorndike, Edward L. (1940). 37
Stone, Clarence R. (1941). 38
Thorndike, Edward L. (1941) 39
Dale, Edgar (1941) 40
Dolch, Edward W. (1942, 2 studies) 40
Stone, Clarence R. (1942) 42
Rinsland, Henry D. (1945) 43
Hildreth, Gertrude (1948, 2 studies) 43
Kirkpatrick, J. J. and E. E. Cureton (1949). 44
Hughes, Marie M. and V. Cox (1949) 45

MORE RECENT STUDIES: POST-1950. 47

The 1950's

Gentry, Lillian (1950) 47
Kyte, George C. (1953) 48
Kyte, G. C. and V. Neel (1953) 49
Dale, Edgar (1956) 50
Calderon, C. I. (1956) 50
Fry, Edward (1957) 51
Reeves, Olive R. (1958). 53
Fitzgerald, James (1958) 54

The 1960's

Fry, Edward (1960) 54
Dolch, Edward W. (1960). 55
Fullmer, D. W. and C. J. Kolson (1961) 56
Denslow, Orrience (1961) 57
Johnson, Glenn R. (1962) 58
Strickland, Ruth G. (1962) 59
Loban, Walter (1963) 60
Fitzgerald, James (1963) 63
Stone, David R. and V. Bartschi (1963) 64
Francis, Nelson W. (1965). 65

Continued: The 1960's — Page

Card, Wm and V. McDavid (1965) 65
Card, Wm and V. McDavid (1966) 66
Stauffer, R. G. (1966) 67
Hanna, P. R. (1966). 68
Kucera, Henry and W. N. Francis (1967) 70

The 1970's

Barnes, James A. and Nancy M. Barnes (1970). 72
Johnson, Dale (1971) 74
critique of Johnson:
Johns, Jerry (1972). 75
Johns, Jerry (1971). 75
Carroll, John B., P. Davies, and
B. Richmond (1971) 76
Barnes, James A. and Nancy M. Barnes (1972). 78
Durr, William K. (1973). 82
Hillerich, Robert L. (1974). 82
Lowe, A. J. and John Follman (1974). 84
Johns, Jerry (1975). 84

SUMMARY. 87

BIBLIOGRAPHY 90

VOCABULARY STUDIES: SUMMARIZED, REVIEWED, CRITIQUED, AND OFFERED IN AN HISTORICAL PERSPECTIVE

Since the turn of the twentieth century, many "new" and different word lists, both derived and original, were compiled by researchers, many of whom combined and/or compared two or more previously published word lists. Some studies were word frequency counts while others were linguristic studies. Some studies were reading studies, and other studies were spelling, writing, and/or spoken vocabulary studies. Some studies were reported in detail while other studies were sketchy in nature.

This paper presents summaries, reviews, and critiques of vocabulary studies, and is presented in chonological order of publication to offer an historical perspective. (There is only one variance from this format: Published rebuttals and/or critiques are presented immediately following the "subjects" or original studies.)

This paper is divided into two parts. Part one presents the "Early Studies: Pre-1950" and contains the earlier studies which were published before 1950. Part two presents "More Recent Studies: Post-1950" and contains the more recent studies, published since 1950. There is no attempt to classify studies, except in historical sequence. Basically, the studies presented are concerned with the reading, writing, and spoken vocabulary of young children, grades one, two, and three. Also, "original" and primary sources were used, not secondary.

EARLY STUDIES: PRE-1950

The 1910's

One of the earliest studies of the twentieth century was an attempt by Housh (1918) to measure certain qualitative aspects of ten second-grade readers. He assumed that repetition of vocabulary was an important factor in determining the worth of a word. He (1) investigated the entire vocabulary of all ten books; (2) determined the common vocabulary; (3) compared the vocabulary of the "method" and "content" readers ("method" and "content" were not actually defined); and (4) described the relationships of similarity and frequency of use between the two kinds of readers. This study set a pattern that was followed for many years.

The seven content readers were the Riverside Second Reader (1911 edition), the Elson Primary School Reader, Book II (1913 edition), the Cyr Reader, Book II (1901 edition), the Gordon Reader, Second Book (1910 edition), the New Education Reader, Book II (1900 edition), the Baldwin and Bender Reader (1911 edition), and the Heath Second Reader (1903 edition).

Housh tabulated each word in each book. The ten readers had a total of 1,566 pages with a total of 143,789 words. Although he listed the words common to all ten readers and indicated the frequency for each word, he omitted words from his list that had a frequency of less than fourteen. Although he included words with inflected endings, he did not

explain why they should be included as new words.

Housh felt that a critical analysis of frequency of use is important and should be done when selecting basic and/or supplementary reading textbooks.

He noted that only 419 words were common to all ten readers, and the "method" and "content" readers had 655 to 926 words in common. He further noted that some words occurred only two or three times, and he criticized the publishers for failing to develop opportunities for repetition of these words.

The 1920's

Three years later, Packer (1921) examined the content of ten similar second readers to determine the vocabulary burden, as determined by frequency or repetition of words. He examined the following series: Aldin, Beacon, Brooks, Carrol and Brooks, Cyr, Heath, New Education, New National Riverside, and Wheeler. He then systematically compiled a frequency vocabulary list which, because it was not published at that time, was later published by Ernest Horn.

In the same year, Edward L. Thorndike (1921a) published his first in a series of three noted vocabulary lists. He completed a frequency count of sections of children's popular book They were: Black Beauty (11,500 total running words), Little Women (13,000 total running words), Treasure Island (13,000 total running words), Scrooge's Christmas (entire book), The Christmas Carol (8,000 total running words), Sleepy Hollow (entire book), Youth's Companion (entire book, except the fine

print and advertisements--25,000 total running words), all fifty-six selections found in Hosic to be the commonest features of school readers (27,000 total running words), ten primers and first readers (entire books, 80,000 total running words), ten second readers (150,000 total running words), ten third readers (283,000 total running words), Book One of the Thorndikes' Arithmetics (entire book, 32,000 total running words), plus thirty-six other different sources. The thirty-six other sources included materials of standard farming, sewing, literature, trades, correspondence (500,000 total running words), cooking, and newspapers (90,000 total running words).

A word-by-word count of the frequency of each word was completed. As contrasted with Housh and Packer, plurals formed by adding "s" were counted under the singular form. Also, plurals derived by changing "y" to "i" and adding "es," adverbs formed by adding "ly," comparatives formed by adding "r" and "er," superlatives formed by adding "est" or "st" and verb forms derived by adding "s," "ed," "d," "n," or "ing" were counted under the primary forms.

After the tabulation was completed, Thorndike then "credited" each word according to the "category" of occurrences: words having a credit-number of 49 or over were found in the first 1,000 words of importance in the list; a credit number of 29 to 48, 1,001 to 2,000; 19 to 28, 2,001 to 3,000; 14 to 18, 3,001 to 4,000; 10 to 13, 4,001 to 5,144; 9, 5,145 to 5,544; 8, 5,545 to 6,047; 7, 6,048 to 6,618; 6, 6,619 to 7,262; 5, 7,263 to 8,145; 4, 8,146 to 9,190; 3, 9,191 to 10,000. Then, in the

case of the 5,000 most important words, the credit number was followed by a second number combined with a letter which indicated in which thousand and in which half thereof the word belonged. Thus, 2a meant that a word was in the first half of the second 1,000.

Thorndike then published an alphabetical list of the 10,000 most frequent words he found in his study. He also indicated the "credit" of each word so teachers, administrators, curriculum directors, etc. could, by using his list, better evaluate text difficulty. He also chose seventy most useful phonograms and categorized appropriately his first 1,000 most frequent words. Teachers, he suggested, could use this list for phonetic drill. He further suggested that his 10,000 word list would be useful in the selection of foreign language texts, high school texts, spelling lists for stenographers, and reading tests. He encouraged researchers to add to and expand his count of a total of 10,000 different words.

In a later critique of Thorndike's 10,000 Word List, Gates (1926a) noted that only 14 percent of Thorndike's sources were from children's literature. This fact he considered challengeable because many subsequent studies were based on his list and was used as a basis for selecting children's materials as well as in the construction of children's materials. Dolch (1948) considered most of Thorndike's source material applicable for educated adults, but not for young children. Dale (1941) suggested that perhaps 230 words of Thorndike's most frequent first 1,000 probably were not known to children entering first grade.

In reference to Thorndike's 10,000 word list (1921a), Dolch (1928b) also questioned Thorndike's assumption that "frequency" and"difficulty" correlate closely. Since there was no established correlation between frequency levels and school grades, Dolch believed that one could object to Thorndike's "ratings."

Dale (1931) was concerned with several possible sources of error in frequency studies, such as Thorndike's. He felt that a frequency count could indicate possibly the more "familiar" words but could not determine difficulty of the reading material. He also cautioned readers about small samplings, about variability of words known by children, and about "homographs," words spelled the same but with multiple meanings. Furthermore, he felt that frequency counts do not sufficiently consider the multiple meanings of words. He further stated that adding derived words with root forms was an error. Another serious source of error, he believed, was the assumption that the measure of the importance of a technical term in general reading can be secured by adding together frequency scores as discovered in a variety of different sources, thus giving appropriate weights or credits to ranges of occurrences.

In a further reference to Thorndike's 10,000 List (1921a), Gates (1926a) believed that the Thorndike List was the most valid published word list. He pointed out, however, that the study was based on word forms, not meanings. He stated that it was based on 86 percent adult material and only 14 percent children's material. Although the words were arranged in order of frequency, Gates felt that the order of frequency was not necessarily the order in which to teach them to children.

Selke and Selke (1922) investigated the content of twelve beginning books. Each book reportedly had a "distinct method," although "method" was not defined. ("Beginning books" were not defined either, although it is assumed that Selke and Selke examined twelve pre-primers.)

The investigators found 1,636 different words, 38 of which were common to all twelve books. They counted all plurals formed by "s" with the root word. All other derivatives were counted as separate words. Only 783 of the 1,636 words were found in one book, which is only 47 percent of the total number of words in the study.

The investigators reported that 70 percent of the total number of words occurred less than ten times each in four books, 50 to 59 percent in three books, and less than 50 percent in one book. They concluded that there was little agreement in practice as to the number of words a beginning book in reading should introduce. There was a very limited number of words common to all the beginning books in the study and a large number of words appeared only once in any book.

Selke and Selke further stated that introductory reading books cannot be supplementary to each other because different books may introduce words of equal difficulty, but only to a small degree will the words be the same. They also noted that most beginning books contain many words whose frequency is very low.

As chairman of the National Committee on Reading, Gray (1925) reported on a "newly combined" list. The three

combined previously published lists were: (1) the Ernest Horn List of the spoken vocabulary of eighty one-to-six-year-olds (1925); (2) Mrs. Ernest Horn's List of 200,000 running words of Iowan Kindergarten children (1927); and (3) the P.C. Packer List (cited by Gray, 1925) of 70,000 running words of the spoken vocabulary of Detroit first graders.

Results of the combining of these three lists were 5,000 different words which were common to (1) two out of three lists with a frequency of twenty-five or more, or (2) in all three lists with a frequency of fifteen or more.

In another vocabulary study, H. W. Kircher analyzed thirty-seven primers and first readers. Gray (1925) requested and was granted permission to publish Kircher's List, because it was a "later study." (No information concerning procedure, etc. was included in Gray's article.)

In another study, Gates (1926a) compared the Thorndike (1921a) and the G. Dewey Lists (cited by Gates, 1926a) and concluded that the Dewey List added nothing to the Thorndike List. Gates then compared the Packer List (1921) and the Thorndike List (1921a). He found 115 words on the Packer List were not in the first 2,500 of the Thorndike List. Also, 362 words of the first 1,000 of Thorndike's List were not on the Packer List. Gates concluded that it was desirable to give some weight to the choice and order of words used in primary school readers.

Gates then discussed several criteria considered in arranging the vocabulary for primary reading. He stated that

both utility for children and utility for adult reading could be realized. He felt that educators were still not in the position to say that one or a certain number of criteria were more important than other criteria. The material should be of interest to the child and be repetitious in nature to facilitate learning. Length of words was excluded as a criteria because of Gates' observations and because of Horn's rating of the frequency of words. Based on all of the above criteria, Gates decided to study the Thorndike List (the first 2,500 words). He combined derivatives and had a frequency composite computed for each word. He found that there were few new word-forms but there was a new arrangement of the words. Then a judgment was made as to whether or not a word should be taught in the first two years of school. The final status for each word was then determined by the factors listed below:

> 25 percent by judgment of interest to children.
> 25 percent by judgment of utility to children.
> 15 percent by frequency of appearance in primary literature.
> 12½ percent by frequency of appearance in primary readers.
> 12½ percent by frequency of appearance in primary children's spoken language.
> 10 percent by frequency of appearance in adult literature. (Gates, 1926a:636)

Gates then divided his words into three groups, the first, second, and third 500. Within each group the words were classified under the parts of speech in alphabetical order. Gates then suggested several uses for this list. It could be used (1) in the writing of new material, (2) in subjects other than reading, and (3) in the construction of test materials.

In a subsequent study, Gates (1926b) used four sources to obtain 1,500 words which he felt were suitable for use in all forms of reading materials in grades one, two, and three.

He took the 2,500 words of highest frequency from Thorndike's 10,000 Word List (1921a) and counted words in certain selections of children's literature (Moore, Annie, unpublished study as cited by Gates, 1926b) which were not in the 2,500 words from Thorndike's first 1,000 most frequent words. Gates then examined all the words and tabulated the frequencies in a series of readers used by Packer (1921) in his study of primary grades. Any additional words not already included on his growing list were added. Then all additional words found in the 1,000 most frequent words in the spoken vocabulary of young children (Horn, 1925) were added to the previously published Gates List (1926).

He then listed the words under eight parts of speech according to ten criteria. This list of words was evaluated by several "experts" whom he reported were already familiar with the factors known to contribute to difficulty in learning. Each word was then given a numerical rating on six different bases to determine interest of words for children. Interests were connected to facts, activities, and situations of interest at the primary level.

The three "experts" then carefully studied the words and ratings and tabulated composite scores for all of the words. These were then ranked from 1 to 1,500 with the words of highest rank first. Gates then suggested that the words of higher rank should form the sequence in which the words

should be taught.

In a critique of the 1926 Gates List, Wheeler and Howell (1930) found what they termed "curious inconsistencies." Of the four sources of the Gates List, only two were primary sources. Gates also broke up his list into twenty-four separate alphabetical lists. He also classified his list into eight parts of speech. It was, therefore, possible to find the same word in nine different locations.

Working that same year but independently of Gates, Ernest Horn (1926) analyzed several previously published studies to determine the 10,000 most commonly used words in adult writing. His purpose was to provide a spelling list, the source of which was writing outside of school. He also summarized and critically evaluated the various investigations which were utilized in determining the need for such a list. He further desired (1) to discuss the most important problems and techniques involved in this type of vocabulary study and (2) to show how this list may be used not only for practical but also scientific purposes.

He examined sixty-five types of writing with a total of 5,136,816 running words. He had problems in the tabulation because of the various exclusions in the different studies. He used Thorndike's creditation method. He excluded proper names, days of the week and months, words with less than four letters, and forty-one "small" words (not listed). Inflected ending words were counted as different words. He published

the list in alphabetical order with creditation noted after each word.

His sources were: W. E. Chancellor's 1,000 Spelling Words, published in 1910; L. P. Ayers' Spelling Vocabulary of Personal and Business Letters, published in 1913; Anne Nicholson's Speller for the Use of the Teachers of California, published in 1914; Cook and O'Shea's The Child and His Spelling, published in 1914; W. N. Andersen's Spelling Vocabulary List, published in 1921; J. D. Houser's Economics Class Vocabulary List, published in 1916-17; W. F. Clark's Writing Vocabulary List, published in 1921; Ernest Horn's Banker's Letters Vocabulary List, published in 1923; and Ernest Horn's Highly Personal Letters List, published in 1922, as cited in Ernest Horn (1926).

In an often cited study, Madeline Horn (1927; 1928), as Chairman of the Child Study Committee of the International Kindergarten Union, attempted to determine the spoken vocabulary that children used while attending kindergarten. The list of words was obtained by recording verbatim kindergarten children's conversations (1) while attending class; (2) when stimulated by pictures; and (3) used at home. The frequencies of each word were tabulated. A "word" was considered as such if it appeared in Webster's International Dictionary, 1925 edition. Proper nouns were checked in the Biographical Dictionary and in A Pronouncing Gazetter. Two or more words which represented one concept, such as "bean bag" were

counted as one word. Children's words, such as "choo," proper nouns, such as "Humpty Dumpty," slang, such as "gee" and commercial words, such as "jello" were also counted. Colloquialisms, such as "mhm," contractions, and inflections of nouns, verbs, pronouns, and adjectives were also counted as separate words.

The results showed a total of 489,555 running words and 7,097 different words.

The Committee report (1928) listed the words in alphabetical order with word frequencies and "placement" of each word, as Thorndike did. [For example, 1b5. The first number indicates the thousand in which it appears. The letter indicates the first 500 (a = first 500, b = second 500), and the last number, if any, indicates the hundred of the 500 in which the word appears.]

Dolch (1927) combined fifteen previously published lists which included adult and children's writings, speech and reading matter. He wished to establish a graded vocabulary list, grades one through eight.

Dolch used Ernest Horn's Commonest Words in the Spoken Vocabulary of Children up to and Including Six Years of Age List, published in 1925. This list included words gathered from children's speech, and had a total of 1,820 different words of highest frequency out of 5,000 words of different forms found. In addition, he included his own study (1927) called the Free Association Study of Children's

Vocabularies. In that study, children in grades two to eight wrote all the words that they could think of in fifteen minutes. Also included, as reported by Dolch, were (1) W. F. Jones' (1915) investigation of English spelling in children's experimental stories, grades two to eight; (2) children's compositions, grades three to nine, as examined by W. F. Tidyman (1921); (3) Ernest Horn's Basic Writing Vocabulary List (1926); (4) Kircher's List, published in 1925; and (5) Gates' List (1926b). Packer's List (1921) and Thorndike's 10,000 List (1921a) were also used in Dolch's "Combined List" study.

His results indicated that only 24 words were common to all fifteen lists; 4,141 words appeared only once; 4,529 appeared two, three or four times; and 3,935 appeared five times or more. Dolch omitted all duplications of words and proper names from the list. He combined inflectional ending forms. His total number of different words was 12,605.

Even though Dolch's Combined Word List combined many thousands of words from fifteen words lists which dealt with both adult and children's spoken and written vocabulary, and reading materials, Dolch felt that more research was still needed. His list was not complete, he felt.

Dolch (1928b) stated that vocabulary difficulty may cripple understanding and destroy interest. For school children's books there should be interest, understanding, and ease of reading. Vocabulary difficulty is usually related to ease of reading. In this article, Dolch discussed six items to consider when determining vocabulary difficulty.

First, there is a rough formula (total running words divided by the number of different words) which expresses a ratio and gives a percentage. The smaller the percentage, the greater the ease of reading because of more repetition. This formula does not distinguish between the relative differences of different words or multiple meanings. Second, Dolch suggested that his Combined Word List could be used to evaluate reading materials by determining the percentage of words not on his list; however, this does not take total running words into consideration. Therefore, the ratio previously mentioned above might be more accurate. Third, it is necessary to get figures on repetition of difficult words before one can fully understand the word situation in any reader. Fourth, attempt must be made to uncover all the possible facts concerning the qualities of the different books under evaluation so far as word analysis can reveal, in order to get a total picture. Fifth, comparison of the difficult words to the total number of running words must be made. Sixth, there is a need to calculate the median frequency of each grade level group of difficult words in order to arrive at a more accurate picture of the distribution of difficult words in a series of graded texts.

Dolch wrote that supplementary readers should be read with little help. He believed that studies indicated that children choose supplementary reading materials two or three grade levels lower than the child's actual grade placement. Therefore, teachers should analyze vocabulary

difficulty of supplementary texts, keeping the findings of research in mind.

In addition, Dolch (1928a) discussed his objections to sampling procedures, and he stated that texts are not and can not be homogeneous.

> They must include a series of stories, acticles, or topics which differ a great deal from one another. Therefore no section of a book can fairly represent all sections (1928:171).

Dolch stated that if one samples lines at regular intervals throughout, then it is not a piece of reading material that is being evaluated but a succession of "disconnected bits." "In fact, the conditions of language and style in writing are such that no study of pieces can truly represent the whole (1928a)."

He also pointed out that when considering the difficulty of textbooks (grades one to six) that texts on the same grade level were often of varying difficulties, some easier, some harder and above "grade level" in over-all difficulty.

With reference to vocabulary studies, Ethel Fennell (1928) wrote that texts have repetitions of words but certain words have entirely different meanings. She felt that this is a bigger problem than most people realize. She examined fourteen readers, from primers through third grade which were published between 1918 and 1926. She listed reoccurring words whose meanings differed. She constructed oral reading exercises on this list and administered the exercises in

first and second grade classrooms in two New Jersey schools. She noted that the children were "mixed" racially with average I.Q.'s. Results of her small study indicated that the children experienced "serious difficulty" with her tests. (Examples of the tests were not given.)

Fennell concluded that a teacher should be aware of the problem and should, therefore, stress the teaching of word meanings.

In a critique of Fennell's study, Hockett (1937) reported that Fennell found only nineteen to thirty-three words occurred with different meanings in five pages. But Hockett felt that these meanings should not prove difficult for a child because these are meanings with which he is already familiar, as indicated by previous spoken vocabulary studies.

The next year, Sister Irmina (1929) compared the vocabulary content of primary readers commonly used in Catholic schools with the Gates Basic Vocabulary List (1926b) and the public school readers. She concluded that Catholic readers are similar in vocabulary content to public school readers, with the Catholic readers having a higher percentage of words common to the Gates List. She concurred with Dolch that there is a general lack of repetition of words, both within a book and from book to book. She also indicated that the Gates List was a more reliable criteria for the evaluation of reading textbooks.

Then Selke (1929) analyzed the vocabulary of ten spellers, published between 1923 and 1926, to see if there was a close agreement in vocabularies and in grade placement of words. All words were counted and grade placement of each word was noted. All proper nouns, proper adjectives, abbreviations, contractions, possessives, and the word "I" were not counted. Homographs were counted under common spellings. There were 8,427 total different words, with 1,080 words common to all ten spelling books. Only 2,350 words (28 percent of all words) were found in only one speller. Only three words were located in all ten spellers on the same grade level. His conclusion was that there was little, if any, real agreement between the ten spellers in grade placement.

The 1930's

In one of the most frequently quoted studies of the 1930's, Wheeler and Howell (1930) investigated how close the Gates List (1926b) corresponded with twenty recently published readers (ten primers and ten first readers) published between 1922 and 1930. Each page was checked. The total number of words, the frequencies, and the sum of the frequencies were tabulated. All variants, except plurals with "s," were counted separately. Later, however, variants were combined to compare with the Gates List. Proper names were not tabulated. Then the authors made a list of 2,219 different words from a total of 131,000 running words. Raw

frequency of each word was multiplied by the number of books in which the word appeared to obtain a "total frequency." Then 453 words with the highest "total" frequencies were ranked, and both lists were compared. The first one hundred in both lists had sixty-eight words in common. The authors concluded that when evaluating reading materials, one should (1) use the Kindergarten Union List (1928), (2) eliminate words used less than ten times, and (3) compare with the Wheeler-Howell List (1930).

That same year Dolch (1930) presented an additional critique of sampling techniques. He wrote that material derived from sampling is unrepresentative in character. It is almost, if not entirely, impossible to make a statistical adjustment, in order to get data to represent the whole book. The exception is, of course, unless the reading material is very repetitious and is written in a uniform presentation. He tabulated three different samplings from a book, tabulating every tenth page, beginning at different places in the book. He then tabulated the entire book. The three samplings were consistently "higher" than the actual total tabulation. Therefore, Dolch questioned the extent to which any vocabulary data based on sampling is representative of the whole unit of reading material.

In this article, Dolch also proposed an "Index of Difficulty" as one means of evaluating reading vocabulary

difficulty. Index of difficulty is determined by dividing the number of difficult words into the total number of words in the book. One could interpret this as being so many words per one hundred running words.

That same year, Erich Selke (1930) again studied the vocabularies of twelve beginning books in reading (1) to determine to what extent the "lists" had influenced the vocabularies of beginning books and (2) to compare the findings of this study with his previous study (1922).

Selke counted plurals with the root word. All other derivatives were counted as separate words. Hyphenated words were counted as separate words. Each word in each book was tabulated with its frequency. He found 1,207 different words with 582 words (48 percent) in only one book. This was approximately the same percentage as in Selke's 1922 study. There were fifty-two words in common to all twelve books. This list was checked with the Gates List (1926b), Thorndike List (1921), and the lists presented in the Twenty-fourth Yearbook (Gray, 1925). The Selke List had 281 words not found on the Gates List, 67 not found on the Thorndike List, and 355 words not found on the Twenty-fourth Yearbook Lists.

Selke concluded that there still were too few words in common between the beginning readers. He further stated that too many words appeared only once in a book or with too few repetitions.

The next year Edgar Dale (1931) attempted to evaluate

Thorndike's List (1921a) by constructing tests based on those words. The author administered these tests to children in various grades and tabulated the mean percentages of known words in each successive thousand words.

Dale felt that such a study was necessary because (1) the correlation between the frequency of a word and unfamiliarity is +1.0, (2) the mean familiarity of the words in the fifth thousand is less than the mean familiarity of the words in the fourth thousand, (3) the list of "more familiar" words are of little help to a teacher, and (4) one still cannot really determine the level of difficulty with such a list. Therefore, a mean percentage of children in the various grades who will know the words in Thorndike's various categories is needed to help teachers.

That same year Sidney Harring (1931) attempted to find the words common to fifteen primers. His tabulation indicated a total running words of 77,004 words with 1,260 different words. Only 34 words were common to all fifteen books and 124 words occurred only once. In one book, only 538 of the 1,260 words occurred. Harring then compared his 1,260 words to Thorndike's List (1921a). He found that 220 words were not on Thorndike's List. Also, 66 were not on the entire Thorndike 10,000 List and 344 words were not on the Gates List.

Three years later James Fitgerald (1934) analyzed the vocabulary of children's letters written outside the

school to determine the vocabulary used and spelling errors made.

Elementary school children were asked to give letters (1) received through the mail from young friends and (2) which were not written in school. The total number of letters, written by pupils in grades four, five, and six, was 3,184. There were 742 letters from fourth grade pupils, 1,199 from fifth graders, and 1,243 from sixth graders. Girls wrote 2,269 letters while boys wrote 915 letters. Letters were from forty-one of the states and dated between 1929 and 1930. There were 1,149 letters from rural areas and 2,035 from town and city areas. There was a total of 461,321 running words and 7,340 different words, excluding 145 expressions not found in Webster's New International Dictionary.

Fitzgerald then listed, in alphabetical order, with (1) frequency of use and (2) frequency of errors the 2,106 words which occurred eight times or more. The 2,106 words were 97 percent of the total running words, and 2,000 were used in all three grades. He felt that the list could assist curriculum-makers in selecting words for spelling.

As a follow-up to his 1921 study, Thorndike (1931) analyzed 4½ million words from books recommended for pupils in grades three through eight. He analyzed the differences between the vocabulary of 120 juvenile books and vocabulary of 279 sources used in determining the Thorndike 20,000 Word List. This material was more general in nature. Credits were given to each word, as was done in his 10,00 Word List.

He then analyzed various books recommended for pupils in grades three through eight, noting the frequency of occurrence of the various words in the various books.

That same year Gates (1935) published his revised list of 1,811 words. He had many sources: (1) the 2,500 words of highest frequency from Thorndike's List (1921a) of 4½ million words, 14 percent of which was from children's literature; (2) the words not included on the 1,000 most frequent words on the Thorndike List but on the Anne Moore List (cited by Gates, 1926b); (3) the first 1,000 most frequent words on the Packer List (1921) which analyzed ten first readers; (4) the 1,000 most frequent words on the Ernest Horn List of spoken vocabulary of children (1925); (5) various other studies which he did not list; and (6) other words in at least one sixth of a list of 105 books (readers, supplementary readers, stories, etc.). Results indicated 5,600 different words which were (1) ranked by "experts" in order of interest to children and (2) judged by "experts" in regard to utility for children. The revised list was not divided into parts of speech because it was felt that it provided no great service and was too cumbersome.

This was followed by Edward Dolch (1936a) who compiled the Dolch Basic Sight Vocabulary of 220 Words from words (except nouns) common to the Gates List (1926b), the International Kindergarten Union List (1928), and the Wheeler-Howell List (1930).

Using the 500 words of most frequency from both the Gates List and the International Kindergarten Union List, plus the entire Wheeler-Howell List, Dolch compiled the lists on a dictionary basis (that is, regularly inflected forms of a single root were combined) and a comparison of the three lists were made. Words common to all three lists were chosen. Dolch also included twenty-seven words in his list, even though they were common to only two of the lists. The twenty-seven had high frequencies (Dolch, 1938, 1949) and seemed to go with other words on the list, for example, "go" and "goes." Dolch called his 220 words "basic" because the list included the "tool" or "service" words that were used in all writing, no matter what the subject. The "service" words included such parts of speech as conjunctions, prepositions, adverbs, adjectives, and verbs. Dolch did not include nouns because he felt that the nouns changed with subject matter.

As an extension of this study, Dolch (1939) checked the list against thousand-word samplings (despite his previous critique of sampling procedures) of textbooks in four subjects. The Dolch Words were found to include 70 percent of the running words in first grade books, about 65 percent of the running words in second and third grade books, and about 60 percent of the running words in most books for grades four to six inclusive.

In the same study he indicated that the 220 words

students in fourth grade and many fifth and sixth graders are such poor readers that they do not recognize instantly all of the Dolch 220 Words.

Dolch further commented that a child suspected of having a poor sight vocabulary should be tested on the Dolch Words to see which words he did not know. He should then be trained to recognize instantly by sight the words he did not know.

Two years later, Dolch (1941) published the following table which shows the percentage of the Dolch Basic Sight Vocabulary of 220 Words as compared to the total running words in school texts in four subjects.

Percentages* of Dolch Basic Sight Vocabulary of 220 Words in the Content Areas

Subject	Number of series	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Reading	4	70	66	65	61	59	59
Arithmetic	2			62	63	57	57
Geography	2				60	59	54
History	2				57	53	52

*Based on a 1,000 word sampling of each book, including inflected forms of the basic sight words in which the sight word appears unchanged.

Dolch also wrote (1948) that almost one half of the words in any book, magazine, or newspaper is made up of the Dolch Basic Sight Vocabulary of 220 Words. He did not claim that the 220 words included all the words that a pupil should know. He felt that a pupil should at least know the 220 basic words, and he stated that an important use of the list is in remedial work. He felt that the remedial process would be complete if a child is taught on sight his unknown "instant" words plus some sounding help, and "context" guessing. The pupil then should be able to do the learning from books that school work demands.

To prove his point, Dolch cited a study he conducted (1936b) when he used seventy-five fourth grade students who knew 194 or less of the Dolch Basic Sight Vocabulary of 220 Words. Each Child has his own unknown "instant" words in a manila envelope. Each day, the child put the cards into two stacks, one known, one unknown. The children were administered the Gates Silent Reading Test (Type B--Reading to Predict Outcome of Given Events), both pre-tests and post-tests. Results indicated an average gain of two months after one month of remedial work. Dolch concluded that sight vocabulary deficiencies can be corrected in a short period of time and correction can be made with a large group under normal classroom conditions.

The Dolch Basic Sight Vocabulary of 220 Words

a	done*	I	out	these
about	don't	if	over	they
after	down	in	own*	think
again	draw	into	pick*	this
all	drink	is	play	those*
always*	eat	it	please	three
am	eight*	its*	pretty	to
an	every	jump	pull	today
and	fall	just	put	together*
any	far	keep	ran	too
are	fast	kind	read	try*
around	find	know	red	two
as	first	laugh	ride	under
ask	five*	let	right	up
at	fly	light*	round	upon*
ate	for	like	run	us
away	found	little	said	use*
be	four	live	saw	very
because	from	long	say	walk
been	full	look	see	want
before*	funny	made	seven*	warm
best*	gave	make	shall	was
better*	get	many	she	wash*
big	give	may	show	we
black	go	me	sing	well*
blue	goes*	much	sit	went
both	going	must	six*	were
bring	good	my	sleep	what
brown	got	myself*	small	when
but	green	never	so	where
buy	grow	new	some	which*
by	had	no	soon	white
call	has	not	start*	who
came	have	now	stop	why
can	he	of	take	will
carry	help	off	tell	wish
clean*	her	old	ten	with
cold	here	on	thank	work
come	him	once	that	would
could	his	one	the	write*
cut	hold	only	their	yellow
did	hot	open	them	yes
do	how	or	then	you
does	hurt*	our	there	your

*Words common to only two out of the three lists that Dolch added to the list because they were of high frequency.

In a later critique of Dolch's mandate, Clarence Stone (1950) questioned the accuracy of Dolch's conclusions and doubted the advisability of the omission of nouns from his basic sight word list. He suggested that even if a child knows, as a result of drill, all 220 Dolch Words, that the child may still have an inadequate reading vocabulary.

An additional study during this time was made by Hocket (1936). He compared the vocabularies of thirty-three primers which were published between 1923 and 1935. He analyzed each page, tabulated the frequency of each word, and counted all variants as separate words, except "s" which he counted with the root word. Hyphenated words were counted with the root word, unless very simple. Then the separate parts were counted as different words. Title page, table of contents, preface, and the concluding word list were not included in the tabulation; however, the comprehension exercises, poems, and titles of stories were included in the count.

Hockett found 1,713 total number of different words, 174,076 total running words in all thirty-three books, 25 words common to all thirty-three books, and 63 words common to thirty of the thirty-three books. The median number of running words per book was 5,094 words and the average was 5,275. The average vocabulary burden was 303 words and the median was 287. Only 40 percent (681 words) appeared in one book.

In a subsequent and similar study of Dolch's 1927 study, Buckingham and Dolch (1936) combined eleven previously published lists to derive the "Combined Word List." The researchers felt that a graded vocabulary was indeed important. The eleven sources were: The International Kindergarten Union List (1928); Gates' Free Association List (1926b); Jones' List (1915); Tidyman'e List (1921); Studley and Ware's Essentials in Spelling List, published in 1914 (as cited by Buckingham and Dolch, 1936); Payne-Garrison Spelling List, published in 1931 (as cited by Buckingham and Dolch, 1936); the New Orleans Public School Spelling List, published in 1916 (as cited by Buckingham and Dolch, 1936); Ernest Horn's Basic Writing List (1926); Gates' Reading Vocabulary for Primary Grades List (1926b); and Thorndike's 20,000 Word List (1931).

Buckinghamd and Dolch followed Thorndike's rules for tabulation which he followed when preparing his 20,000 Word List. Proper names, abbreviations, contractions and archaic and poetic forms were not tabulated. Any words found on the International Kindergarten Union List were not tabulated as far as the frequency count was concerned. The authors felt that those words had already been proven important and had high frequencies.

To determine the size of each grade's vocabulary, the authors used two criteria: (1) the vocabulary development of children, and (2) experience with children. For grade one, the International Kindergarten Union List was used. For grade two, the IKU List and 984 other words with a

frequency of three or more were listed. It was felt that one half of the grade two words probably belonged in grade one. For the other grades (three to eight), words not already placed on previous grades' lists were carried forward to the next grade's list, using words with less and less frequencies. They found 2,481 words occurred more than once. The authors felt that the list was indeed incomplete, had gaps in the grade levels, and did not emphasize meanings. They published the list in alphabetical order with a letter(s) after each word to indicate the source from which it came. A number indicated the frequency by the thousand.

Another vocabulary study was made by C. T. Gray (1936) who reported on a derived word list which combined several other word lists and which categorized 19,000 words. Included were the Thorndike List (1921a), the Gates List (1926b), the Horn List (1925), plus eleven different other lists, including the Dolch "Free-Association" study (1937).

Then McKee (1937) challenged frequency lists that may say they determine difficulty through frequency counts. McKee wrote,

> ... The fundamental measure of the difficulty of a printed word is the degree of familiarity which the reader has with the concept or meaning that the printed symbol represents in the setting in which it is used. (1937:242)

Although McKee agreed with Ernest Horn that using word lists in an uncritical and mechanical fashion is not educationally sound, he did think that the Thorndike List (1921a) at least contributed to preventing wide disparity in

the choice of words for children's books.

That same year Hockett (1937) analyzed the vocabularies of twenty-eight first grade readers. He wanted to present information that would make it easier for teachers to select appropriate books for students. He used the vocabularies of thirteen readers published before 1930 and fifteen readers published between 1930 and 1935. Each reader was analyzed and tabulation was made of the number of times each word was repeated. All variants were counted as new words, except plurals formed by "s". Words assumed to appear at least forty times were not tabulated. They were "at," "all," "and," "he," "I," "is," "said," "the," "to," and "you." The title page, table of contents, preface, and concluding word lists were omitted in making the analysis because they were not normally read by the children. Comprehension exercises, poems, and titles were counted. The workbook exercises were not included. The total number of words were counted and totaled on every second page. An adding machine was then used to total the running words for every book.

As compared to his previous studies, Hockett felt that (1) there was a definite trend toward smaller vocabularies, (2) the percentage of words on the Gates List (1935) of 1,811 words increased 5.4 percent, (3) the average repetition increased 3.0 words, and (4) there was a reduction of 16 percent of the vocabulary load by more than 100 words. In the average (five out of eight books) first reader, 64 percent

of the words were in the Gates first 500. In five out of six first readers, 83 percent of the words were in the Gates first 1,000. About 7.4 percent fell outside the Gates List. The median book length was 8,540 words. The average book length was 9,057 words. The average number of different words was 589 and the median was 581.

Hockett (1938) continued his investigation as he analyzed twenty-nine second grade readers. He tabulated each page, and he recorded every different word found in the book. He then used these different words to make what he called a skeleton list. Then he went through all of the books, page by page, and completed a frequency count of each word, in each book. Any word not on the skeleton list was written in at the appropriate place. Commonly derived forms were included with the root forms. After the list for each book was completed, the list was compared to the Gates List (1935). The total running words for each page was counted and written at the bottom of the page. An adding machine was used to total the total running words. Table of contents, the concluding word lists in the back of the book, and all explanatory material directed to the teacher were included in the tabulation.

The typical second grade reader had an average of 1,000 different words, and a total of 21,000 running words. The second grade reader typically was two times as long as an average first reader and four times as long as an average

primer. About 52 percent of the Hockett List was in the Gates first 500 most frequent words. Hockett observed a reduction of vocabulary load in the more recent readers. In the pre-1930 readers, the average number of different words was 1,146. In the 1930 to 1933 readers, there was an average of 1,057 different words, which amounted to an 8 percent reduction. In the readers published between 1934 and 1937, there was an average number of 913 different words, which was a further reduction of 14 percent.

Then Fitzgerald (1938), with a desire to supplement his studies at the fifth and sixth grade levels, studied the vocabulary and spelling errors of third grade children's life-letters. He requested real letters which were received through the mail and were written by third graders. From twenty-seven states, 1,256 letters were collected, 539 from boys and 717 from girls. About 1,000 letters came from cities and towns and 200 came from rural areas. There were 100,840 total running words, 2,928 different words, and 8,504 tabulated spelling errors. Fitzgerald then listed 692 words which were used ten times or more with frequency of use and frequency of error counts. The 100 most common words were used 69,191 times throughout the letters.

In another vocabulary study during this period of time, Betts (1939) reported a study of vocabularies of first grade books in order to establish a basic reading materials list. Betts tabulated (1) the total number of pages at each

level, (2) the total number of different words at each level, (3) the total number of running words at each level, (4) the number of words on a page at each level, (5) the number of words common to all basal readers at each level, and (6) the number of words common to a given number of readers at each level. He studied the nature of the words common to a given number of books at each level, and the vocabulary control as evidenced by repetition or frequency. He attempted to determine the consistency of words introduced on one level.

The readers used were published between 1932 and 1937, and the books were from thirteen different series. Reading readiness or complementary books, covers, inside fly leaves, preface or introductions, table of contents, directions or notes to the teacher, concluding words lists, words within the illustrations, and numbers of pages were not tabulated. Title of story units, title within the units, including both print and artist's lettering, all story content, and comprehension exercises were counted. Betts reported his findings in two ways: (1) the basic list of words at each of the five levels, and (2) a basic cumulative list for the primary grades. The second list shows not only the frequency for each key word but also the frequency and spread of the primary form and all of the variants. Compound words, hyphenated words, contractions, and abbreviations were counted as separate words. Each word was counted on the basis of spelling, not meaning.

Betts found a total of 210,283 running words from thirty-nine first grade readers, with 12,155 from pre-primers, 69,969 from primers, and 128,159 from first readers. The average number of running words in a pre-primer was 935, 5,382 in a primer, and 9,858 in a first reader. Three key words were common to thirteen pre-primers, "a," "the," and "mother." Thirty-four words were common to thirteen primers and three of these common to the pre-primers also. There were 166 words common to all the first readers.

Betts concluded that the tabulation of all forms of key words only at the primary level did not provide sufficient data on vocabulary. Data on frequency and spread of word frequency appeared important. There was a wide disparity in the vocabularies of the books. He suggested that the data from this study could help the teacher to check against spelling vocabularies and vocabulary overlap in the readers. (Betts did not include his two lists in his report.)

Fielstra and Curtis (1939) compared the two Thorndike Lists, the 10,000 (1921a) and the 20,000 (1931) Lists, to determine to what extent the words found in the first 5,000 words levels of the 10,000 List fail to appear in the corresponding 1,000 word levels of the 20,000 List. They attempted to discover to what extent the words appearing on the original list appear on the 1,000 word levels above the tenth in the later list. They also wondered to what extent were words occurring in the first 10,000 Word List not in the original list.

About 4.52 percent (226 words) of the first 10,000 List were found in different 1,000 word levels in the 20,000 List. Results indicated more numberous shifts in positions of words in the upper half of the 10,000 List to the other 1,000 word levels in the 20,000 List. About 95 percent (215 words) of the "displaced words" on the 10,000 List found lower positions. About 12.19 percent (1,219 words) of the 10,000 List were not among the first 10,000 words on the 20,000 List. About 12.10 percent (1,210 words) of the first half of the 20,000 List and 29.1 percent or 1,167 words (almost half) of all the words in the seventh, eighth, and ninth 1,000 word levels of the 20,000 List do not appear on the 10,000 List. The authors of the study concluded that the 20,000 List was not just a continuation of the 10,000 List, it was a definite revision. Results of any vocabulary study based on the 10,000 List would differ markedly if based on the 20,000 List (the first 10,000 words only). However, if the study was based only upon the first 6,000 words of both lists, and if one did not particularly care with which thousands one was working, then there would be no problem. The authors cast doubts upon the validity of results of studies based upon the 10,000 List.

The 1940's

As chairman of the committee of the National Conference of Research in English formed to review current research in vocabulary, Seegers (1940) reported briefly, without any table or lists, on the recent research of both derived and

original lists. Generalizations were made that sounded just like the philosophies of Gates, Dolch, Horn and Thorndike. Then the committee reported that the most important studies which were carefully compiled, were that of Ernest Horn, Madeline Horn, Buckingham, Dolch, McKee and McKee, Fitzgerald, Tidyman, Lorge, Gates, and Thorndike. These lists yielded the most important information and the limiting factors were carefully stated. Rinsland's List (1945) had not been published yet, but was mentioned as another important study.

In a stinging critique of the committee's report, Thorndike (1940) said that, "It is regrettable that so many of the makers of counts have left so much of their data unpublished." (The authors of this paper found this to be quite true. Because of this problem, often some procedural data, some results, and even some actual word lists of some of the more important authors could not be reported; however, this by no means implies that if an author's list is not included in this paper that the author failed to publish it.)

Thorndike made several suggestions to the committee. They should have specified to which Thorndike List they referred. The committee did not include several studies which he felt were indeed very important. Thorndike felt that the committee could not discount unusual discrepancies in small counts of 2,000 to 200,000. He felt that this could be done with larger counts of 2,000,000 words or more. He also felt that the makers of counts should not include slang, contractions, proper names, nouns, pronouns, places,

and numbers (for example, 5,6, or 9). He questioned whether compound words should be counted as separate words or if derivatives should be counted as separate words. He felt that there was a lack of consideration of the general psychology and linguistic factors in methods used to obtain the lists of words. Homographs, or words spelled the same but with different meanings, presented a problem, too. Thorndike warned against the idea that one list must be learned more than another. Grade placement of the words was, he felt, a poor idea because some teachers, ". . . unprivileged in respect to intellect . . ." might teach only words on her grade level. Thorndike suggested that publishers should compare the new texts to the first 2,500 words of the Thorndike List (1937) and supply such an analysis to the buyers. Thorndike felt that progress had been made in the field, but that more research was needed.

In an attempt to provide current data, Clarence Stone (1941) compared the vocabulary of twenty pre-primers which were published between 1930 and 1940. He did not tabulate words for a frequency count. He did list the words that appeared in each book and counted the total number of different words and found 359 words. Only 186 appeared in only one book, 54 appeared in two books and 29 appeared in three books. He found 100 words that were most widely used. Ninety words appeared in four or more books. Stone quoted from a frequency count (Hayward and Ordway, 1937) which did both frequency and

range of use of 211 words in fifteen pre-primers. All the words used by Stone had a high frequency in Hayward and Ordway's study.

Stone also assumed that if a word appeared in four or more pre-primers, it would have a high frequency of use. Also, if a word appeared in only one or two pre-primers, it would not be an important word, even if it had a high frequency.

Stone further felt that the revised Gates List (1935) of 1,811 words did not include simpler derivatives and, therefore, Stone's List could be contrasted with the Gates List, because he felt simpler words tend to appear in pre-primers.

Then Stone compared his 100 most important words to the Gates List. Five pre-primers did not meet Stone's expectations, but nine did. (He felt that 75 percent of all the pre-primer's words should be on the Gates List.) Stone further felt that, to be fair, one should compare the average number of repetitions for each book based on the same number of running words in each book. Longer books obviously would have more repetitions; therefore, Stone felt that repetition alone is a poor gauge of difficulty.

In his third extensive study, Thorndike (1944) added 10,000 more words to his 1921 and 1931 Word Lists to make his 30,000 Word List. Again, he included inflected ending words under the root word. Thorndike then presented a new design which was different from his 1921 and 1931 designs in

reporting his word list. He suggested to the teachers the grade levels at which the words might be taught. He stated that the list could be used for non-English speaking people, as well as good readers, depending on needs.

Despite the admonition of previous research, Edgar Dale (1941) compared the International Kindergarten Union List (1928) and the most frequent 1,000 words from Thorndike's 10,000 Word List (1921a). Dale listed all words common to the two lists, noting which words appeared in the second 500 of the Thorndike List. He then noted that 230 words were not in Thorndike's first 1,000 and suggested that possibly these 230 were not known to children entering first grade.

Dale wrote further that the use of vocabulary must meet two criteria: (1) the vocabulary must be known by children and (2) the vocabulary must be of permanent value to children in reading activities. He issued two cautions in the interpretations of his list: (1) the list contained a large number of homographs and (2) it cannot be assumed that all words on the Kindergarten Union List will be known to all first grade children.

This list is still known and used as the Stone 769 Easy Word List. It is also included in the Spache Readability formula for primary grades.

In still another study, Dolch (1942) attempted to determine those words which made up most of the words written by the average person ("average" was not defined.) Dolch

used the Gates List (1926a) of 2,500 words which was shortened by comparing it with the following word lists: (1) the Fitzgerald List (cited by Dolch, 1942) of 2,106 words which was written by children in grades four, five, and six; (2) the Smith List (cited by Dolch, 1942) of 2,156 words, which was written on school papers by pupils in grades two through eight; and (3) the Dolch Free Association List of 9,520 words (1936). Words with less evidence of common use either by grade placement or frequency were dropped. The final list contained 2,000 commonest words for spelling. According to Dolch, these words were so selected to include words which made up to 95 percent of all the words written by the "average" person.

In another study in the same year, Dolch (1942) reported his analysis of four studies which included (1) Carl Wise's 13,641 words common to twenty spellers (cited by Dolch, 1942); (2) L. P. Ayres' first 1,000 Spelling Words (cited by Dolch, 1942) as found in 5,000 children's compositions; (3) Tidyman's 2,000 Word List (1921) from children's compositions; and (4) the International Kindergarten Union List (1928) of the spoken vocabulary of children entering first grade. Dolch concluded that any list of 2,000 words or more includes words that are seldom used by children, for example,

some of the words on the Kindergarten Union List were used only by seven children. Dolch wrote that truly common words are few in number.

In a reappraisal and extension of studies of the vocabulary of the previous decade, Stone (1942) studied 107 primary books published between the years 1930 and 1941. His analysis was of twenty-one series from pre-primer to third readers. The 5,314 count included 286 names and 846 derivatives ("s," "es," "d," "ing," "n," "en," "y," "ie," "ly," "er," "est," "y" to "i" and then add "es," "less," "ied," "ed," "ier," and "ily"). Stone then assembled each word into particular categories of levels of introduction. For example, if a word appeared in seven or more of the twenty-one pre-primers, then it was included on the pre-primer list. If a word appeared less than three times in the third grade readers, then it was put on the fourth grade reader level. Altogether, there were ten levels of words. The result was the formulation of an up-to-date "graded" vocabulary for the primary grades. Stone also noted each word that appeared on the Kindergarten Union List (1928), the Gates List (1935), the Buckingham-Dolch Combined Word List (1936), Berlund's Fourth Grade Vocabulary (cited by Stone, 1942), and Durrell's Vocabulary for Fourth Grade (cited by Stone, 1942). Stone concluded that the trend of increasing the number of pre-primers was inadequate. We needed, he felt, adequate vocabulary expansion with sufficiently easy material at each level.

In the first major study reported after World War II, Rinsland (1945) analyzed 1 percent of twenty million elementary school children's uncorrected, freest writings. From a total of 6,102,359 running words, he reported the raw frequencies of words from 100,212 papers from children in all [illegible] from each child. Rinsland himself felt that the sampling from the first and second grades was inadequate. Many of the papers were very short and others were incomplete or unfinished; therefore, he included in his study the spoken vocabulary of first graders as reported by Fry, published in 1931 (cited by Rinsland, 1945), and Trent, published in 1931 (cited by Rinsland, 1945).

Plurals, contractions, abbreviations, and inflectives were counted as separate words, as was in the Horn study. Most slang, trade names, baby talk, and made-up words were not included. Rinsland then indicated the position of frequency in the same manner as did Thorndike. His published list of 14,571 words included words which occurred three or more times in any grade, one through eight. He compared his results with those of previous research and felt that there was little agreement among the authors as to the vocabulary introduced. He noted that seventeen authors agreed on only one word in total words and in grade placement. The word was "long."

In two separate studies Gertrude Hildreth (1948a, 1948b) sampled 19 percent of Rinsland's List (1945) and concluded that only 2,800 words "do the most work." She indicated that only a small proportion of the words were used

frequently and there was a wide range of rarely used words that accounted for many a reading disability case.

In her second study, she compared Dale's 769 Easy Words (1941), the Dolch Basic Sight Vocabulary of 220 Words (1936a), and the Rinsland Word List (1945). She observed a large amount of over-lapping plus some important differences. She found a total of only 156 words common to all three lists, and she recommended the use of the "new" list (Dale, Dolch and Rinsland Lists combined) to teachers:

> These are the words that all children should certainly learn to spell and to recognize in print during the elementary school years, and they would ordinarily take precedence in drill over less common words. Any word that appears in all three lists rates priority in language instruction. If a child knows these basic words he has a frame-work for learning less commonly used words (Hildreth, 1948b:41).

Hildreth further felt that (1) it would be wrong to drill in isolation and (2) primary grade teaching should not entail teaching <u>all</u> of these words before other words. Words should be taught in meaningful context. However, she suggested teachers should keep an eye on this list when preparing spelling and reading lists.

In a related study of item difficulty, Kirkpatrick and Cureton (1949) compared four methods of determining vocabulary burden: (1) criteria based on frequency, (2) criteria based on judgment, (3) criteria based on syllable count, and (4) criteria based on frequency and judgment.

The authors used Thorndike's frequency groups for this study. In the judgment portion of the study, five judges were asked to rearrange eleven groups of cards in

order of frequencies. The syllable count included the key word of each of the items. The results of this study indicated that frequency plus judgement, as evaluated by comparison to the Thorndike List, appeared best. The syllable count appeared the least desirable.

In a pioneer study, Hughes and Cox (1949) investigated the relationships of language and vocabulary of first grade children to their reading textbooks. The authors desired to answer some important questions: What were the differences in vocabulary used by children and by books? What differences could be found in the richness and vitality of expression? What differences were there in the maturity of sentences as measured by length and number of relational words? What concepts and interests are employed in speech and not utilized or expanded in the beginning books?

Hughes and Cox recorded the free conversation of two classes of Detroit first grade students during the "show-and-tell" for a period of two months. The median I.Q. score was 101.67, based on the Detroit Beginning First Grade Test, Form A. They then analyzed ten pre-primers and six primers. The books were (1) W. Gray's Curriculum Foundation Series, (2) G. Hildreth's Easy Growth in Reading Series, (3) M. O'Donnel's Alice and Jerry Series, and N. B. Smith's Learning to Read Series.

Results indicated that 331 words common to the books and to children's spoken vocabulary. There were 1,097 different words spoken by children and 401 different words in the

books. The sentences were more complicated in the children's speech. Their sentences were more "dynamic" and had a greater number of relationships. The authors suggested that teachers use experience stories, reading charts, and booklets to better reflect the real speech patterns employed by pupils.

MORE RECENT STUDIES: POST-1950

While the pre-1950 studies were mainly concerned with an analysis of vocabulary through frequency word counts and comparison of vocabulary lists of previously published lists, the last 3 decades, the 1950's to the present, gave evidence of some change, with a shift in emphasis to a linguistic analysis of the language. The linguistic analysis studies concerned themselves with the effect of an unstructured vocabulary. There were, however, continued replications of the studies conducted in the twenties, thirties, and forties.

In the first study published in the fifties, Gentry (1950) constructed a list of common words for beginning readers and listed the pre-primers which contained the smallest number of different words but had a large percentage of common words. Sixty-six pre-primers with copyright dates from 1930 to 1948 were used. The total number of different words were tabulated. The number of books in which the word appeared was recorded. The study produced a list of 600 words. Two hundred fifty of the 600 words were considered to be "common words" (words common to three or more of the pre-primers). The 250 words were categorized into two groups. The first 125 words appeared in nine or more of the total number of pre-primers. The second 125 words appeared in less

than nine of the pre-primers. The two groups of words were used to rank each pre-primer. The pre-primers whose total vocabulary consisted of the highest percentage of the "first 125 words" was ranked first. The vocabulary load in nine of the pre-primers was found to be 100 percent of the "first 125 words" list. Only 38 percent of the vocabulary in the lowest ranking book could be found in the first half of the "common words." Eighty-five percent of the vocabulary in the median ranking books could be found in the first half of the "common words." The percentage of words in a given pre-primer outside the "common words" list ranged from a high of 41 percent to a low of 3 percent.

In commenting on the use of word counts in relationship to spelling, Aykes and other investigators, wrote Hildreth (1951), discovered an important principle: very few words make up most of the words in writing. "A hundred words take care of over half our writing needs even in adult life, and a thousand words do about 90 percent of the work (Hildreth, 1951:257)."

In an attempt to bring together the commonalities in spelling and reading and to construct a core vocabulary, Kyte (1953) presented a list: (1) to develop efficiency in reading, oral and written expression, and spelling in the elementary school; and (2) to provide the minimal essentials for children, especially for the non-English speaking child

and for the adult illiterate. The "core words" were taken from the first 500 words in the Thorndike-Lorge List (1941) of 30,000 words,and the first 500 of Horn's List (1926a) of commonly written adult words. The 500 most commonly used words in the Rinsland List (1945) of children's writing compared with the "core words." A core vocabulary of 663 words was derived. Three hundred seventy-two words formed a basic list of highly frequent reading and writing words. One hundred twenty-nine words were more important for writing than reading, and the remaining 162 words were more important for reading than writing. The common words from the Rinsland List furnished evidence that the core words were probably familiar to learners having previous background in school.

In another similar attempt at a "core" list, Kyte and Neel (1953) wanted to provide the minimal essentials for various instructional programs. They constructed a core vocabulary of spelling words using the Horn List (1926) for adult writing and Rinsland's Basic Writing List (1945) used by elementary children. The words from both lists were selected according to the following criteria: (1) words which occurred most commonly in adult and in children's writings; (2) words which occurred most frequently in adult writing but less commonly used in children's writings; and (3) words which occurred most commonly in children's writing but less commonly in adult writing. The basic list consisted of 501 words. The core words were categorized into five

types: Type one contained words which appeared in identical form among the 500 most commonly used words in adult writing and in children's writing. Type two contained fifty-five words among the 300 most frequently used words in children's writing. Type three contained twenty-one words occurring in the fourth 100 in children's writing and in the second 500 of adult's writing. Type four contained thirty-eight words occurring in the first 500 of adults' writing and in the second 500 of children's writing and also in the first 500 in one elementary school grade. Type five contained twenty-seven words occurring in the first 500 of adults' writing and in the second 500 in at least one elementary school grade. Type six contained forty-five words occurring in the first 330 adults' writing but not in the first 1,000 in any elementary school grade.

In an article concerning the problem of vocabulary in reading, Dale cited Horn as follows:

> Two thousand words with their repetitions make up 85.05 percent of the running words, in adult writing; 3,000, 96.9 percent; 4,000 97.8 percent; and 10,000, 99.4 percent.
>
> A limited number of words--1,000 or 2,000, certainly not more than 3,000--give us the easily predicted words. From then on we have moved behond the words of common experience--the structural words in the English language, the household words, the words of time, place, directions, parts of the body, and the like. (Horn, as cited by Dale, 1956:114)

In a slight shift of emphasis and concerns, Calderon (1956) investigated several units of study (1) to determine

the minimum vocabulary that the greatest percentage of educators would agree upon and (2) to provide the basis for a sufficient English vocabulary for the child who speaks little or no English. Six lists were selected for study: three from Texas, two from California, and one from New Mexico. The list contained units of study concerning saftey, foods, home, social, environment, etc. Each word as well as the name of the list was printed on a card. Words that appeared on all six lists were considered to be words that 100 percent of the educators would agree upon. Words that appeared on five lists, four lists, three lists, two lists, and one list were considered to be words having 85 percent, 66 percent, and 50 percent, 33 percent, and 16 percent, respectively, on which educators would agree. A total of 381 words were found to have 50 to 100 percent agreement. Nineteen additional words considered important to the study were added.

The next year Fry (1957) developed a list of the most frequently appearing words for the remedial reading

teacher or classroom teacher. Several criteria were used to develop the word list: (1) to select the most frequently used words and (2) to edit the words to exclude easily recognized variants and nouns of limited use. Two standards were used to achieve the criteria: (1) several word counts containing the most frequent words and (2) personal experience as a remedial teacher as well as subjective logic. For example, the word list included the words one through ten in its first 500 words but excluded nine. Fry added nine. Fry omitted babyish sounding words like candy and daddy for the sake of high school teachers. The word counts used were the first 500 of Thorndike-Lorge 30,000 Words (1941), the Rinsland List (1945) of children's writing, and the Faucett List (cited by Fry, 1957) which combines the most frequent words on the Thorndike Count and the Horn List (cited by Fry, 1957). Further reference was made to the Fitzpatrick List (cited by Fry, 1957) and the Dolch List of 2,000 Commonest Words for Spelling (cited by Fry, 1957). The study produced a list of 600 "Instant Words." The list was divided into twenty-four groups of twenty-five words each. Groups smaller than twenty-five were not possible because of the way the words were presented in the word counts. Fry concluded:

> The first part of the list is much more valuable and accurate than the latter part. That is, the first group of 25 words is definitely more frequently used than the second group of 25, the first 100 more frequent than the second hundred; the first half (300 words) more frequent than the second half. There is high agreement on all the scientific word counts for the first part of the list.

> Conversely, it is doubtful if the 23rd group of 25 words is much more frequently used than the 24th group or even that some of these words should be included in a list of 600 most common English words. But let me hasten to add that almost all the words on this list, including the words in group 24 (the last group of 25 words) appear in either the Thorndike-Lorge first 500 or Rinsland's first 600, and usually on several other lists also. (1957:456)

In an attempt to replicate earlier studies of vocabulary of basic readers, Reeves (1958) conducted a study of the vocabulary introduced in seven beginning reading series. He analyzed the pre-primers, primers, and first readers of Scott, Foresman; Macmillan; Ginn; Houghton-Mifflin; Winston; Row, Peterson; and Allyn and Bacon. (The publishing dates were not indicated.) All the words introduced in each of the books were tabulated. Proper names of characters, pets, and toys were omitted. Root words and their inflected forms were counted as one word. The result was a total of 633 different words. One hundred nine of the 633 words were common to all seven series. Whether they were common in grade was not mentioned by the author.

Forty-one additional words appeared in six of the seven series. The words introduced in the basic pre-primers of all the series totaled 115. Fifty-seven or 66 percent of these words made up the 109 common to all seven series, but there were 231 of the 633 different words which appeared in only one of the series. The number of words appearing in only one series varied from four in one series to sixty-four in another series.

In an endeavor to determine what words children should study in spelling, Fitzgerald (1958) selected 449 from his list of 2,650. His list of 2,650, which is recorded in his book A Basic Life Spelling Vocabulary, was obtained by determining the overlap of child and adult writing among the lists of Ernest Horn (1926), Rinsland (1945), McKee-Fitzgerald (cited by Fitzgerald, 1958), and other less extensive investigations. The list purported to make up about 94 percent of the running words written by children and adults throughout their lives. Fitzgerald concluded that the 449 words selected were the most used core words for writing. On the average, they had the highest credit ratings from the vocabularies from which they were derived. These words and their repetitions comprise more than 75 percent of the running words ordinarily written. Furthermore, wrote Fitzgerald:

> Every one of the words were written in children's letters and in school themes. All but three were employed frequently by adults. All were found in third-grade letter writing, and in second-grade compositions. (1958:224)

The 1960's

Another study by Fry (1960) found that the 300 most common words of his earlier study (1957) made up an average of 63 percent of the words in the reading texts of three major publishers in grades one through three. The range varied from 58 to 77 percent. These same 300 "Instant Words" made up nearly one half of most adult reading material which

included newspapers, magazines, and popular books. "It seldom dips below 40 percent, even in technical articles," wrote Fry (1960:38). Many of these "Instant Words" do not adhere to patterns or phonic rules. He suggested that mastery of these words may not be realistically expected until the third grade level of reading.

Dolch (1960) reported that when he attempted to write children's stories with Dolch Words only, other words not in the Dolch List had to be used. These other needed words fell into two categories or special classes: special words (such as would be needed in Dolch's Tiger Story) and needed words (people, relationships, action words). In an attempt to find the words really needed he examined fifteen different books which he had written. These books covered a wide range of subjects ranging from Indians to real-life stories to folk stories. The vocabulary of each book consisted mostly of his 220 Sight Words and his 95 Common Nouns. Extra words were only used when necessary. The total number of words in each of the fifteen books were listed. The list included three kinds of words: the permitted, the special words, and the more general words. From this list Dolch selected the 315 permitted, all of the words used in one book only, and all of the words not on the First Thousand List (no source given). This study produced a list of 684 words which actual experience had shown were needed for simple story telling. These words were found to be "needed extra

words." They were all within the limits of the First Thousand Words for children's reading which in itself was found to be too broad for "hesitant readers."

Fullmer and Kolson (1961) wanted a small, frequently-used word list which would guide the teacher in developing word recognition necessary for success in beginning reading. Words from eleven basal reading series were used. A total of forty-five pre-primers, primers, and first readers were studied. All the published basic reading series between 1954-1959 were included: Allyn and Bacon; American Book Company; Ginn and Company; D. C. Heath; Houghton-Mifflin; Lyons and Carnagan; the Macmillan Company; Row, Peterson Company; Scott, Foresman; and World Book Company. Root words and their derivations were counted as different words. The words were divided into five lists: List I included words appearing in at least ten of the series; List II were the words appearing in nine of the series; List III were the words appearing in eight of the series; List IV were the words appearing in seven of the series; and List V were the words appearing in six or less of the series. List V was dropped because of its low count. The four list were correlated with the International Kindergarten Union List (1928) and the Dolch Basic Sight Vocabulary (1936a). A list of 184 words was obtained. It was considered small because it was smaller than the widely used Dolch List. The words were considered frequent since over half of the words were Dolch Words. The

list met the needs of Fulmer in the following ways: (1) the level of application was for first grade; (2) it was manageable since the words were broken down into four lists; (3) its utilitarian function was suggested for testing readiness for formal reading instruction, for assessing sight vocabulary deficiencies, and for establishing reading groups; (4) the teacher could use the list as a guide in preparing experience charts and other reading materials; (5) the list had been derived through consistent word count techniques used by Thorndike and others; (6) the reading series used were current; (7) the words were significant since all the words were listed in the International Kindergarten Union List and 64 percent of the words were on the Dolch List.

Denslow (1961) checked the word frequency and the reading difficulty in eight first grade science books. The books were published during 1947 to 1961 by Scott, Foresman; Heath; Macmillan; Winston; Ginn; Singer; Lyons and Carnahan; and Scribner. The words selected were listed in at least six of the eight books and occurred four or more times in each book. Words were considered "new" if pluralized by adding other than s, if used as a possessive, and if variant endings were added. Compound words were counted as two words. Numerals were counted as words. Proper names of characters, pets, and toys were not counted. Sentences were selected and counted. The Dolch List and Gates List were used to determine the difficulty of the words. The Spache Readability Formula was used to measure the vocabulary and the sentence

structure. Only 61 of the 451 words that occurred in all sight science books were found to appear in six or more of the texts. Also, the publishers' grade designation was found not to present an accurate representation of the vocabulary burden or sentence complexity.

In another replication of earlier designs, Johnson (1962) examined seven basal reading series (1) to determine if there was an overlap of vocabulary between series, and (2) to compile a core vocabulary for grades one through six. The seven basic reading series were American Book, 1949; Ginn and Company; 1953; Houghton, Mifflin Company, 1957; Lyons and Carnahan, 1949; Macmillan Company, 1951; Row, Peterson and Company, 1947; and Scott, Foresman and Company, 1955. The "core" consisted of root words, proper nouns, compound words, hyphenated words, and onomatopoeic words. All words that appeared in five or more series were listed as core words. Johnson, as a result of his data, limited his core words to the first three grades since so few words met the core word criteria in the intermediate textbooks. Very little vocabulary overlap was found between the series. The following recommendations were offered:

> 1. Teachers use the core list to help children develop common words they will come in contact with in their various reading experiences.
> 2. Teachers use the core list as a method of evaluating the difficulty of various books (....this would be but one method of evaluating books ...).
> 3. The core list could be used by teachers in developing charts, teacher-made tests, in teaching opposites and similarities, and in teaching prefixes and suffixes.

4. The core list could be used as an additional chick list in teaching spelling, (Johnson, 1962:471).

In an attempt to analyze the structure of children's language in the first through sixth grades, following somewhat the pattern of the Hughes and Cox pioneer attempt in 1949, Strickland (1962) compared the aural-oral language of children with language patterns contained in the books the children were expected to read. She also attempted to detect at a selected grade level, the influence of any apparent differences on the quality of children's reading skill.

The spoken language of first through sixth grade pupils was recorded. The spoken language was analyzed in the following areas: syntactic structure of sentences, the frequency of occurrences of certain patterns of syntax, the amount and kinds of subordination, length of sentences, and flow of language. The relationships of age, sex, intelligence of the children and socioeconomic and educational levels of their parents were studied. The textbooks which the children studied were analyzed to determine at what point the language patterns freely used by children began to appear in books. Representative samples were taken from each reader to determine the occurrences of patterns used by children. At the sixth grade, the quality of children's spoken language was compared with that of their silent reading comprehension, oral reading interpretation, and their listening comprehension. The number of children selected by random sampling was 575 children from sixteen public schools

in Bloomington, Indiana. The sample included 100 children in each of the six grades with the one exception of the fourth grade. Only seventy-five children were studied in grade four.

Results of the study indicated that children used a large number of patterns of structure. Some patterns occurred more frequently than others. The most used patterns occurred frequently at all grade levels. The lengths of the phonological units used by children varied more within a grade than from grade to grade. The basic subject-verb-object pattern was the only one used basically in all the books sampled. The pattern differed from book to book in a given series as well as from series to series. Sentence patterns appeared to be introduced at random in a rather haphazard manner. Also, there were no given patterns repeated or controlled for the purpose of mastery or readability. Sixth grade children who ranked high in silent reading comprehension, oral reading comprehension, and listening comprehension made more use of the common structural patterns than did children who ranked low on these variables. The higher ranking children used fewer short sentences and had a higher mean sentence length.

In an attempt to study the language of children in kindergarten through grade six, Loban (1963) attempted (1) to find out if there were predictable stages of language growth; (2) to identify a definite sequence in language development;

(3) to develop methods of analysis to aid the scientific study of children's language; and (4) to determine how children vary in ability with language and gain proficiency in using it.

Loban selected a representative group of 338 kindergarten children in 1952. At regular intervals for eleven years, samples of their language were selected. Two subgroups were formed from the representative group: a group of thirty subjects exceptionally high in language ability and a group of twenty-four subjects exceptionally low in language ability. The language samples consisted of reading, writing, speaking, and listening. Data were collected by (1) new methods of analysis which were the outcome of a 1959 conference on linguistics sponsored by the United States Department of Health, Education and Welfare, and (2) by methods based on previous research, tests, and ratings or indices.

The new analysis method of combining meaningful syntactic units with phonological methods of segmentation, and by identifying and dealing with the noncommunicative elements, the mazes, proved to be a new and useful measure of language. The following language finding were obtained from this study :

1. Children in the first seven years of school spoke more words, increased their number of communicative units, and increased the number of words spoken in each of

those communication units. The high subgroup used more words and units than did the low subgroup. Even so, the high subgroup reduced their number of units they needed for expression through subordination.

2. The subjects as a whole decreased the number of mazes and words in mazes. However, the average number of words in mazes increased for the low subgroup.

3. The same number of words among the 12,000 most commonly used words (from Thorndike's The Teacher's Word Book of 30,000 Words, 1944) were used by the low and high subgroups of subjects. The low subgroup used more of the next 20,000 words (from the 13,000 to 33,000); whereas, the high subgroup used more of the least commonly used words above 33,000.

4. The high group was significantly more fluent than the random group, but their readiness of response did not differ from the random sample. In contrast, the low subgroup's fluency and response was less fluent and slower in response than the random sample.

5. The differences in structural patterns used by the two subgroups were negligible except for the linking verb pattern and the use of partials (incomplete units of communication).

6. Although pattern differences (small elements used within structures) show considerable differences between the high and low subgroups, the high group used a large repertoire of clauses and multiples (movables within movables).

7. Lack of agreement between subjects and predicates and consistency with verb tense proved to be the most frequent kind of deviation.

8. Students high in language ability were also high in reading ability. Students low in language ability were low in reading ability.

9. Writing ability was related to socioeconomic positions. Those who were in the four lowest socioeconomic categories were below average in writing. Those who were highest were above average in writing.

10. Adverbs and noun clauses were used more frequently than adjective clauses.

11. The subjects who were low in writing were low in reading achievement. Those who were high in writing were high in reading achievement. Similar patterns were found in those with high and low in oral proficiency.

12. There was a low positive relationship between health and language proficiency.

Fitzgerald (1963) developed an "intergrated" core vocabulary for listening, speaking, reading, writing, spelling, and handwriting which he suggested as useful in communication as well as vocabulary drvelopment. Five hundred commonest words of Madeline Horn's Kindergarten List (1928) were compared with the most frequently used 500 of the Gates Vocabulary for Primary Reading (1935). The frequency counts for each of the words were obtained from each

of the following sources: The I.K.U. List (1928), The Gates Primary Reading Vocabulary (1935), The McKee-Fitzgerald Child Letter-Writing Vocabulary (unpublished, as cited in Fitzgerald,1963), Rinsland's Elementary School Composition Vocabulary (1945), Dolch Basic Sight Vocabulary (1936a), Ernest Horn's Adult Writing Vocabulary (1926), and the Thorndike-Lorge Comprehensive Reading Vocabulary (cited by Fitzgerald,1963). The study produced a total vocabulary of 644 words.

Stone and Bartschi (1963) compiled a composite list of words introduced in five most widely used basal reading series together with the Dolch 2,000 Words for Better Spelling (1942 and Fry's 300 Instant Words (1960). The series used were Scott, Foresman and Company; Ginn and Company; Macmillan Company; Houghton, Mifflin Company; and D. C. Heath and Company. The percentage of the Dolch 2,000 Words within the total words of each basal reader was determined. Also, each word was identified as a word from either the first half of the Dolch Basic Word List, the second half of the Dolch Basic Word List, the first 100 Fry Instant Word List, or the second 100 Fry Instant Word List. The composite list was categorized in one of the following grade levels: 1^1, 1^2, 1^3, 2^1, 2^2, 3^1, and 3^2 by taking each word's mean grade level placement as indicated in the basals, and two word lists. They found that the "basic" and "instant" words were almost completed by the beginning of the second half of second grade. The total words introduced in each of the five series for

grades one through three contained the following percentage of words from the Dolch 2,000 Words: Scott, Foresman, 74 percent; Ginn, 67 percent; Macmillan, 73 percent; Houghton-Mifflin, 68 percent; and Heath, 43 percent.

Francis (1965:267) prepared a corpus containing over a million running words of present-day edited English. The materials came from the Brown University Library, the Providence Athenaeum, New York Public Library, and one of the largest secondhand magazine stores in New York City. The material was selected from fifteen main categories, four of which were "Press," "Religion," "Learned," and "General Fiction." The samples and selected pages were chosen by a random sampling method. Five hundred samples of approximately two thousand words were taken from each of the categories. A data processing system was used, and word frequencies were taken from one million running words. (See Kucera-Frances, 1967).

In two related studies, Card and McDavid (1965) compared the relative frequency of words in children's writing with the frequency of the same words in adult writing for the purpose of gaining some insight into children's language. The 501 most frequently used words from Rinsland's A Basic Vocabulary of Elementary School Children (1945) were used. The 500 most common words in Horn's Basic Writing Vocabulary (adult writing compiled during 1916-23), and the words of highest frequency in Dewey's Relative Frequency of English Speech Sounds (adult reading, writing, speaking sources collected during 1918), were used as reference points. The

rank orders of the words under study were compared rather than their raw frequencies. Unlike the other two lists, Horn's words were given a "credit number" which was partly based on frequency and partly on how many different kinds of letters the given word appeared in; therefore, the rank order was derived from the credit number. The structured words under consideration were categorized in the appropriate category: verb forms, modal auxiliaries, familial terms, determiners and demonstratives, numerative adjectives, prepositions or adverbials, subordinators, and conjunctions.

Card and McDavid (1966) wanted to demonstrate that there were peculiarities or biases inherent in corporas of English. One hundred twenty-two words in the first 285,062 running words from George K. Monroe's dissertation (as cited by Francis, 1965) were listed in order of frequency. The rank order of these words contained in three other word counts were compared to the rank order in the Monroe List. The three word counts were Godfrey Dewey's Relative Frequency of English Speech Sounds (a composite taken in 1918 of newspapers, magazines, fiction, and drama), Miles L. Hanley's Word Index to James Joyce's Ulysses which was composed during the years 1914-1929, and Rinsland's List of Children's Writing collected in 1937. The examiners concluded that the bias of the corpora are much more apparent when three or four lists were compared rather than two. The farther down the

rank order one goes the greater the discrepancies in the rank orders. Corpuses containing largely expository samples will show bias when compared to corpuses that are largely fictional in content.

An examination of primary textbook vocabularies was directed by Stauffer (1966). He included in his study seven graded reading series published between 1945 and 1950 in addition to three series in arithmetic, three series in health, and three series in science with copyright dates ranging from 1944 to 1954. The "new words" were taken from a list of words in the back of each basal reader. The "new words" from the content texts were obtained by counting each page. Word variants, contractions, and compound words were counted as "new words." A master word list was made for the basal readers and for each of the content areas. It was suggested that the results be interpreted with caution since the textbooks were of earlier dates; however, the assumption was made that the vocabularies were not too different from those found in more recent publications. Only 117 of the 570 first grade words were common to all seven series. Only 7 of the 2,155 words introduced in the third grade level were common to all seven series. The data suggested that the words used in different reading series do not overlap but became more and more different after first grade. In fact, Stauffer wrote: ". . . if a pupil were exposed to all different words in seven reading series (analyzed by Stauffer), he

would be prepared to deal with only about half the words presented in the three content areas studied." (1966:146) The words common to the seven reading series and to the arithmetic series were Level 1--and, are, away, happy, I, is, make, one, see, the, two; Level 2--only; Level 3--instead. Similarly, the words common to the seven reading series and to the science series were Level 1--a, did, do, find, for, home, it, look, make, on, she, the, we, what, will, with; Level 2--(none); Level 3--teeth. Stauffer concluded individualized reading programs offer an opportunity to fill the vocabulary gap among series. The primary vocabulary alone in basic readers and content areas can be estimated to total 6,000 words. Word attack skills "focus on meaning as the principal need with phonetic-structural attack for pronunciation purposes as auxiliary skills . . . is essential if the child is going to become an independent reader." (Stauffer, 1966:146)

In an attempt to help improve the school spelling program, Hanna (1966) attempted to conduct an analysis of the degree and the characteristics of the correspondences that exist between the spoken and written language. Hanna posed two questions: Does the American-English orthography approximate the alphabetic principle? What are the relationships between the phonological structure of the spoken language and its representations?

A total of 17,310 words were selected from two

sources: (1) Thorndike-Lorge's The Teacher's Word Book of 30,000 Words (1944), Part I. Part I contained 19,440 entries. These entries were decreased to 15,284 entries in order to exclude proper names, foreign words, slang, contracted word forms, etc. And, (2) Merriam-Webster's New Collegiate Dictionary, sixth edition, furnished 2,026 "common core" words not listed in Part I of the Thorndike-Lorge List. Each listing was analyzed and the data computerized. The listings were described and interpreted in the following ways:

> 1. An analysis of phoneme-grapheme correspondences irrespective of other phonological factors.
> 2. An analysis of phoneme-grapheme correspondences to position in syllables.
> 3. An analysis of phoneme-grapheme correspondences in relation to their position in stressed and unstressed syllables. (Hanna, 1966:15)

Results indicated that only six of twenty-two vowel phonemes equaled or exceeded 80 percent concordance with the alphabetic (or phonemic writing) principle. These vowels were primarily the short vowel phonemes and their occurrences before the letter "r." The sixteen vowel phonemes falling below the 80 percent criterion were the long vowel phonemes, and the long vowel phonemes before "r." Nineteen of the thirty consonant phonemes were equal to or exceeded the 80 percent criterion. Fifty-two phonemes approximated the alphabetic principle 73.13 percent of the time.

> Graphemic options do not equally distribute themselves throughout a given position. Thus, when the distribution of graphemic option is tabulated for a given position, in many instances a particular graphemic option will exceed the 80 percent criterion in either initial, medial, or final position in syllables. (Hanna, 1966:81)

Thus, instead of only six of the twenty-two vowel phonemes equaling or exceeding 80 percent approximately to the alphabetic principle, an additional nine vowel phonemes by position in syllables have at least one graphemic option occurring over 80 percent of the time. As for consonant phonemes, an additional two more exceeded the 80 percent criterion when the initial, medial, and final syllable position is considered.

It was concluded that American English is primarily an alphabetical language. The average consistency with the alphabetical language is 73 percent when a given phoneme is considered. The predictability of graphemic options in a syllable increases the consistency to 79 percent. The factor of stress added to position increases the grapheme predictability to over 84 percent.

In one of the most reputable computerized studies of the century, Lucera and Francis (1967) presented a collection of lexical and statistical data of the Standard Corpus of Present-Day Edited American English, a computer-processible corpus of language texts assembled at Brown University during the 1963-1964 school year.

Lucera and Francis head four major objectives:

(1) to present the lexical and statistical data about the corpus, (2) to offer useful material for the development and improvement of statistical procedures of linguistic analyses; (3) to make possible the construction

of more satisfactory mathematical models of language; and (4) to compile a corpus of printed English. The authors did not wish to test the validity of various mathematical models of language. The authors, also, did not wish to tabulate a "basic sight" or "most common" words list.

The corpus contains 1,014,232 words, of "natural-language text," fully described in the Manual of Information. The manual and copies of the computer magnetic tape of the Corpus are available, at a nominal cost (Providence: Department of Linguistics, Brown University, 1964). The Corpus, divided in 500 samples, of approximately 2,000 words each, was "synchronized." Only data to texts published in the 1961 was utilized. Only material printed in the United States, with no more than 50 percent dialogue, was included in the study. Footnotes, tables, and picture captions were omitted. The 500 samples were distributed among fifteen categories of subject matter and prose styles. The major decisions regarding content of texts, etc. were made in a conference held at Brown University of J. Carroll, W. Francis, P. Gove, H. Kucera, P. O'Connor, and R. Quick, with procedural help from H. Peyton,Jr., and A. R. Taylor. Several "proofings" were conducted, both before and after keypunching. A computer coding system was utilized, and, therefore, two "versions" of the Corpus were developed: (1) Form A is the full version, which includes most of the punctuation marks, other coding symbols and various "types" of combinations of letters, numbers, symbols,

initials, graphic units, etc.; and (2) Form B is the "stripped" version and excludes most of the punctuation marks and other coding symbols.

In the Corpus, homographs are lumped together as the same type. Variant spellings of phonologically and lexically identical words are listed and counted separately, while syntactically and morphological variant graphic forms of lexically identical words are listed separately.

Lucera-Francis presented the complete Corpus vocabulary (1) in order of frequency of occurrence and (2) in alphabetical order. In addition, a list of the first 100 most frequent words was presented, with the distribution of occurrence of each in the various portions of the Corpus was specified. Word-frequency tables and graphs, plus word-length analysis, sentence-length analyses, and an analysis of the lognormal model were presented and discussed in great detail.

The 1970's

To test the relevancy of the Dolch Basic Sight Vocabulary of 220 Words in more recent literature/ vocabulary studies, Barnes and Barnes (1970) analyzed the occurance of the Dolch 220 in various lists compiled and published in recent years. In Gentry's List (1950) of common words for beginning readers, 88 of the Dolch Words were found in the first half of the 125 "common words." A total of 49 Dolch Words were found in the second half. In the Kyte and Neel List (1953), a total of 190 Dolch Words were found. In Fry's Instant Word

List, (1957), 212 Dolch Words were found. In the Reeves Study (1958), Barnes and Barnes found that (1) 109 words common to the seven series studies contained 93 Dolch Words; (2) the additional forty-one words common to six of the seven series contained twenty-two Dolch Words; and (3) the 115 basic pre-primer words also contained 75 of the Dolch Words.

Barnes and Barnes further reported that 131 Dolch Words appeared in the Fitgerald List (1958), and 117 Dolch Words were in the Fullmer and Kolson List (1961). In the Denslow List (1961), there were forty-seven of the 61 words mentioned by Denslow. In the Johnson study (1962), 137 of the 194 core words found in five or more of the seven first grade readers were Dolch Words, and 23 of the 107 core words found in five or more of the second grade readers were Dolch Words.

Barnes and Barnes further found that the Fitzgerald List (1963) contained 119 of the 220 Dolch Words.

Then, in the Stone and Bartshi (1963) List, 44 Dolch Words were found among the fifty-eight pre-primer words, fifty-five among the 79 primer words, fifty-one among the 115 first reader words, seventeen among the 90 second reader--first book (2^1), ten among the second reader--second book (2^2), five among the third reader--first book (3^1), and none among the 505 third reader-- second book (3^2). A total of 192 Dolch Words were found in this list.

Barnes and Barnes also found 98 of the 122 words of the Francis List (1965) were Dolch Words, using the words used

200 times or more.

Analysis of Dolch 220 Words Found in Other Recently Published Vocabulary Lists

Published Lists	Number of Dolch Words Found In List
Gentry's List (1950)	190
Fry's Instant Words (1957)	212
Fitzgerald List (1958)	131
Fullmer and Kolson List (1961)	117
Johnson Study (1962)	137
Fitzgerald List (1963)	119
Stone and Barschi (1963)	192
Francis List (1965)	122

Barnes and Barnes reported that most of the Dolch Basic Sight Vocabulary of 220 Words were contained in lists provided by researchers of the various, recently published studies, reported above, thus indicating the Dolch Words are indeed relevant and not "outdated." They concluded that the Dolch Words were highly frequent, even in adult reading and writing material, as well as on children's reading, writing, and spoken language.

In a re-examination of the Dolch 220 Word List, Johnson (1971) contended that because the Dolch List was compiled using 1920 data sources, it was outdated. He supported this generalization by comparing the Dolch List to the top 220 words from The Kurcera-Francis Corpus (1967) which was compiled based on adult reading materials. Johnson assumed that if the 220 Dolch Words did not appear in the top 220 of the Kurcera-Francis

Corpus of 50,406 words, then the Dolch List was outdated.

It might be questioned as to why Johnson arbitarily used only the top 220 words from the Kurcera-Francis List, and, further, why he only chose the Kurcera-Francis List. Johnson's findings have not been supported by other recent Dolch comparisons, Johns (1972), Barnes and Barnes (1970, Lowe and Follman (1974), and Hillerich (1974).

In a stinging rebuttal to Johnson (1971), Johns challenged Johnson's contention that the Dolch list had outlived its usefulness. Johns compared the 315 Dolch Combined List to the most frequent in the Hass List, the Kucera-Francis Corpus, and the Wepman List. (See Johns, 1971). In fact Johns stated, the Dolch Words come closer than the Kucera-Francis Corpus words to representing words that are used by young children. In closing, Johns offered several "corrections" of the tables in Johnson's actlcle.

In another analysis of the relevancy of the Dolch 220 Words, Johns (1971) compared the vocabularies of five recently published basic reading series. His purpose was to determine if the Dolch List is still the core of words that comprised 50 to 70 percent of the running words in more recently published basic reading series. He examined all pages in pre-primers and primers and a random sample (300 words) from each one third of each text, grades one through six. A frequency count was made of the Dolch Words.

Percentage of Dolch Basic Sight Words
Found in Five Basic Reading Series

Basic Reading Series	Reader Levels							
	PP	P	1	2	3	4	5	6
Allyn and Bacon	69	70	64	60	52	53	54	54
Ginn	77	64	73	63	53	57	56	58
Lippincott	40	44	56	61	52	49	56	51
Macmillan	55	63	64	57	53	56	54	55
Scott Foresman	79	72	75	65	51	58	56	57

A Comparison of the Percentage of
Dolch Words In Present Study
to Dolch's Original Findings

Investigator	Number of Reading Series	Reader Levels					
		1	2	3	4	5	6
Johns	5	66	61	52	55	55	55
Dolch	4	70	66	65	61	59	59

Johns concluded that there was remarkable agreement between his percentages and those reported by Dolch several decades ago. The slight differences in percentages were noted because of two possibe reasons: (1) Dolch included inflectives and Johns did not and (2) Johns included only reading texts and Dolch included other sources.

In the most extensive, detailed computerized study of the twentieth century, Carrol (1971) examined samples from various published materials to which students are exposed in grades three through nine. The study was designed to produce a

citation base for The American Heritage School Dictionary.

The publications used in the study were named by educator respondents who participated in a national survey of schools in the United States. The survey instrument was a ten page questionnaire mailed to the "highest administrative officers," in 155 public school systems, 44 Roman Catholic diocesan systems, and 22 independent (private) schools. The number of questionnaires used were 71, 11, and 8, respectively. From the 6,162 different nominated titles of materials, kits, novels, poetry, general non-fiction, textbooks, workbooks, etc. to which pupils (grades three through six) are exposed, 1,045 published texts materials were selected. Analysis of the survey responses was conducted by Educational Testing Service. In a computerized analysis, titles were identified, counted, systematically listed, tabulated, categorized, summarized, and placed on a grade-subject matrix. Then 500 word samples of running text were taken, with the actual selection of text samples performed by the American Heritage Dictionary division staff. Headings, captions, footnotes, glossaries, tables, word lists, indexes, teachers' materials, advertisements, phonetic spellings, numbers not in sentences and any other matter not obviously intended for the student were not included in the study. Multiple meanings of words were not considered. Data processing of samples was conducted by the Fullfillment Corporation of America. There were a total of 10,043 samples and a total of 33,623 pages of sampling text.

Two IBM 360 Model 30 computers, with standard card reader and card punch were utilized. There were a total of 5,088,721 running words, with 86,741 different words.

As a result, the American Heritage Intermediate Corpus (AHI) of 86,741 words was developed.

A statistical analysis of the AHI Corpus was conducted, using the "lognormal model," developed by G. Herdan, assuming that the total vocabulary underlying a corpus is distributed according to the "normal distribution" when the logarithms of the frequencies are used. Further, the AHI Corpus was, through computer analysis, compared to several well-known frequency lists, including the Kucera-Francis corpus(1967). The data appeared to justify the general validity of the lognormal model. It was concluded that AHI included mostly common words, with some rarely used words. Also, the higher the grade level, the more difficult and diverse the vocabulary.

The authors presented the AHI in three different manners: (1) in an alphabetical list; (2) in a "rank" order list; and (3) in a "frequency distribution."

It has been noted that a word-frequency count is merely an "experimental attack on our ignorance" to learn something about the structure and composition of a very large, abstract entity by examining a relatively small, concrete part of it.

In a page-by-page frequency count of the Dolch Basic Sight Vocabulary of 220 Words in 32 primary (grades one, two and three), reading, spelling, and English texts, Barnes and

Barnes (1972) examined the relevancy of the Dolch 220, in more recently published primary texts: The Harper and Row Basic Reading Program (1969); The Macmillan Company's Bank Street Readers (1965); Holt, Rinehart and Winston's Sound of Language Series (1966); The Basic Goals in Spelling Series by Kottmeyer and Ware (1967); The Ginn Elementary English Texts for grades one and two (1964); and the Roberts English Series Linguistic Program, Book 3, for grade three. In addition to the above 32 texts "reading" portions of the study, eight of the texts (the spelling and English texts) were used in a page-by-page frequency count of the Dolch words which were to be written directly in the pupils books, in what was termed the "writing" portion of the study.

Barnes and Barnes found that the Dolch 220 words comprised up to 61,78 percent of the 348,904 running words, with 63.37 percent in first grade texts, 60.42 percent in second grade texts, and 59.56 percent in third grade texts.

In the "writing" portion of the study, Barnes and Barnes tabulated a total of 7090 running words and found 39.46 percent were Dolch words.

Most of the Dolch words were introduced by the end of second grade in the spelling and English texts, and by the end of first grade in the Harper and Row and Macmillan Reading Series.

Additional analysis of the date revealed that there were 193 Dolch words which appeared most frequently and comprised about 65 percent of the running words in the primary texts

examined. As a result, the Barnes' Revised Dolch List of 193 Words was presented and it was suggested that the 193 should be taught as "instant" or "sight" words, especially to first grade pupils, pupils with sight vocabulary difficulties, and non-English speaking individuals.

Dolch "Reading" Grade Level Percentages

Grade Level	Running Words	Dolch Words	Percentage of Dolch Words
First Grade	55,406	36,220	65.37
Second Grade	107,610	65,015	60.42
Third Grade	185,888	110,720	59.56
Grades 1-3 Total	348,904	211,955	61.78

Dolch "Writing" Grade Level Percentages

Grade Level	Running Words	Dolch Words	Percentage of Dolch Words
First Grade	504	267	52.98
Second Grade	1,889	997	52.71
Third Grade	4,691	1,534	32.66
Grades 1-3 Totals	7,090	2,798	39.46

THE BARNES' REVISED DOLCH LIST OF 193 WORDS

(in Rank Order of Frequency)

the	did	has	live
to	your	right	under
a	how	their	ride
and	were	after	always
you	down	long	never
in	go	first	run
of	so	where	keep
he	some	again	walk
I	her	find	best
is	from	may	start
said	him	saw	because
that	or	help	kind
it	now	going	fast
was	two	does	its
on	here	made	show
what	know	why	found
for	no	ran	soon
his	come	three	fly
with	look	red	jump
at	good	don't	call
are	too	off	hot
they	make	before	black
do	just	our	those
have	them	eat	gave
not	could	yes	own
all	went	play	cold
one	tell	am	far
we	get	take	today
this	old	away	together
but	big	must	brown
will	if	well	draw
she	these	let	five
write	read	us	bring
can	think	only	round
as	would	small	goes
my	came	better	wish
out	very	get	try
be	new	much	grow
when	say	white	ask
had	put	every	warm
up	many	give	buy
like	which	stop	cut
there	into	work	six
about	who	any	please
then	use	been	funny
see	over	once	ate
little	around	green	
me	want	blue	
by	an	four	

In an effort to "fill a void," Dr. Durr (1973) used library books that primary grade children select to read through free choice. He felt that using library books as a source of data would help "fill a void" in vocabulary studies, which predominately use other, more dictated sources. (For example, a reading text).

Children's librarians in forty various communities were selected, based on varying socioeconomic levels and geographic area. The librarians supplied the popular titles. Durr, for various reasons, did not tabulate the list but presented the list to "experienced" teachers who were "well-acquainted" with children's free reading interests. A final selection of eighty titles was made. A word-by-word frequency count was conducted, through a computer analysis. The total number of running words was 105,280, with 5,791 different words. The "average" word appeared eighteen times. Omitting proper names and onomatopoeic (moo; buzz) and counting only base words of common inflected base words or compound words, the list dwindled to 3,220 different words. He reported the words in descending order, noting the frequencies. He found that 188 words appeared more than eighty-eight times each and comprised 68.41 percent of the running word count. He also stated the 300 most frequent words on the list were what he termed "structure" words.

In a more recent comparison of published vocabulary lists, Hillerich (1974) compared fourteen different lists and

offered still another basic reading/writing vocabulary list: 240 Starter words. The fourteen lists that were compared by Hillerich included, among others, Carroll's List (1971), Hillerich's List (as cited in Hillerich, 1974), the Kucera-Francis corpus (1967), the Rinsland List (1945), the Durr List (1973), the Dolch 220 (1939), the Horn List (1926), the Fitzgerald List (1963), Fry's List (1960), Johns (1972), and Johnson (1971).

Hillerich had two purposes for his vocabulary study: (1) to explore patterns of variation among the word lists and (2) to present a basic, updated vocabulary (his 240 Starter Words) that minimizes the bias of individual counts. In regards to his first purpose, Hillerich reported that variation in word counts was more related to the original source of the words than to the compilation date. He concluded that apparently what authors write for children is different from what children write about.

Regarding compilation dates, the addition of technological words (i.e. space, T.V.) was negligible, and would not present spelling problems.

Hillerich further concluded that the Dolch List did not appear outdated, and the Rinsland List did contain a number of rural and childlike words in little use today.

To compile his 240 Starter Words, Hillerich used the 500 most frequently used words from five different counts, utilizing materials which included the old and the new, juvenile

and adult printed material. From a total of 995 words, Hillerich's 240 Starter Words were identified.

In a rebuttal to Harris, Johnson, and Otto, Lowe and Follman (1974) challenged their implication that the Dolch List had outlived its usefulness. Lowe and Follman compared the Dolch List (the first 150) with: (1) Kucera corpus (compiled from a sampling of adult reading material); (2) the Carroll corpus (compiled from sampling school materials, used in grades three through nine); (3) Otto's list (actually Carroll's third grade list); (4) the D. Johnson list (compiled "children's oral language" from Murphy's list and the first 500 from the Kucera list); (5) the Taylor list (compiled from total word count of basal readers); (6) the Harris list (compiled from total word count of basal readers) and (7) the R. Johnson list.

Lowe and Follman were exacting in their conclusions. The Dolch List is as useful today as it was thirty years ago. There was also a high degree of commonality/similarity between the word lists, derived from a variety of sources, regardless of material, type of material or whether sampling or word counts were made. The basic words were essentially the same.

Lowe and Follman suggested that it is the use of such lists which needs careful attention.

In a more recent frequency count analysis of the Dolch nouns, Johns compared the 95 common nouns to four recently published lists. Johns hoped to derive a short list of nouns which were highly frequent and could be considered as

important as sight words. The four lists used were: (1) The 500 most frequent words from the American Heritage Intermediate (AHI) corpus (Carroll, 1971); (2) the 188 words from Durr's computer study of high frequency words in trade books for children (Durr, 1973); (3) the 500 most frequent words from the Kucera-Francis corpus (Kucera-Francis, 1967); and (4) the 727 words from the 1957 Murphy analysis (as cited in Johns, 1975).

The AHI corpus, compiled from samples of published materials used in grades three through nine, contained 5,088,721 words, from 1045 texts with 500 word samples.

Durr's list was derived from a study of 80 library books popular with primary grade pupils, with 105,280 running words and 5791 different words. One hundred eighty-eight words had more than 88 frequencies and accounted for 68.41 percent of the running words.

The Kucera-Francis corpus contained 1,014,232 words, with 50,406 different words, and was compiled from a sample of varied, published adult reading material.

Murphy analyzed a total of 1,195,098 words in the oral vocabulary of pupils in kindergarten through third grade. He found a total of 6,318 different words, with 727 words used at least fifty times.

Results of the Johns analysis indicated there were 20 nouns common to all four word lists, with 26 others common at least three of the word lists. He then compared these words to the Dolch 95 nouns. He determined that 30 of the nouns also appeared on the Dolch noun's list, as follows:

back	eye	house	school
boy	father	man	table
car	feet	men	thing
children	girl	money	time
day	hand	morning	top
dog	head	mother	tree
door	home	name	water
		night	way

Johns also offered sixteen additional nouns common to at least three of the four lists which did not appear in the Dolch list:

air	friend	nothing	room
book	group	people	side
city	Mr.	place	town
face	Mrs.	road	year

Johns concluded that the forty-six nouns listed above (1) were current; (2) go beyond the original Dolch noun list; (3) are likely to be used in materials read by both children and adults; (4) are worthy of being taught as sight words; and (5) may prove helpful in adult literacy programs.

SUMMARY

The first four decades of the investigations of vocabulary, all utilized the same kinds of techniques. In general, there were no marked improvements from one investigation to another. Each tended to replicate each other; differences existed only in that a study conducted in the twenties was duplicated in the thirties using then current published materials. Other than the Hughes and Cox attempt, there were no major breakthroughs in techniques or approaches. Their influence was negligable until the decade of linguistic influence in the sixties.

As the 1950's approached, the state of vocabulary studies had shown no real improvement. The vocabulary was the same. Publishers still relied heavily upon the same varied vocabulary, with the same large amount of vocabulary, with few exceptions. Only Dolch constantly sought to refine and develop a core vocabulary of the fewest words possible.

The first half century of studies did not appear to influence publishers because texts still were loaded with many different words and fewer repetitions than research suggested.

While earlier studies were mainly concerned with an analysis of vocabulary through frequency word counts and comparison of previously published lists, the last three decades of the 1950's to the present, gave evidence of some

change, a shift in emphasis to a linguistic analysis of the language. The linguistic analysis studies concerned themselves with the effect of an unstructured vocabulary.

There were, however, many replications of the studies conducted in the twenties, thirties, and forties.

Investigators seem to agree that the frequency "order" does not necessarily indicate a sequence in which "sight" or vocabulary words should be taught to children. Several investigators suggested criteria, such as numerical ratings for difficulty (Gates, 1926a). Dale (1941) suggested that the correlation between the frequency of a word and its unfamiliarity is +1.0. McKee (1937) wrote that the difficulty of a word can not be determined by its frequency but instead by its familiarity which the reader has with the concept being represented. Fennell (1928) concluded that although the textbooks may have many repetitions of words, their different meanings may cause more difficulty than publishers might expect. Stone (1941) stated that repetition alone is a poor gauge of difficulty. He felt that one should compare the average number of repetitions for each book based on the same number of running words in each book.

The research indicates that there is little overlap as to the vocabulary used in the various elementary texts, as indicated in the Selke (1930), Harring (1931), Dolch (1927), Hockett (1936), Betts (1939), Stone (1941), Rinsland (1945), Gentry (1950), Calderon (1956), Denslow (1961), Gates (1926a), Johnson (1962), and Barnes and Barnes (1972) studies.

There is sufficient evidence to indicate that few words (less than 3,000 definitely, and perhaps fewer than 1,000) do the most work, as indicated in the Hildreth (1948) Dolch (1942), Dale (1956), Johns,(1971), Durr (1971), Hillerich (1974) and Barnes and Barnes (1970;.1972)studies. Preller (1967) felt that the first 100 most frequent words would be exactly the same regardless of the level of language. Dale (1956), Hildreth (1951), Ayres (as cited in Hildreth, 1951), Fitzgerald (1938), Fry (1957), Dolch (1939;1948), and Barnes and Barnes (1970; 1972) found that from 100 to 449 words make up to 50 to 75 percent of the running words in reading and writing materials. Furthermore, Card and McDavid (1966) implied that the amount of overlap or rank order of words used depends upon the source of material from which the words came. Expository material showed peculiarities when compared to fictional material. In addition, Hillerich (1974) stated that the variation in published word lists is more related to the original source of the words than the dates that the studies were conducted.

BIBLIOGRAPHY

Barnes, James A. Jr., and Nancy Marie Barnes. 1970. An Examination of the Dolch Basic Sight Vocabulary in the California State-Adopted Primary Reading, Primary Spelling, and Primary English Series. Chico, California: California State University, Masters Thesis, Fall.

Barnes, James A. Barnes, Jr., and Nancy Marie Barnes. 1972. "An Examination of the Dolch Basic Sight Vocabulary in California State-Adopted Primary Reading, Primary Spelling, and Primary English Series," California Journal of Educational Research, Vol. 23, No. 4, Sept., pp. 163-176.

Betts, I. A. 1939. "Study of the Vocabularies of First Grade Basal Readers," Elementary English Review, 19:65-69.

Buckingham, B. R. and E. W. Dolch. 1936. A Combined Word List. New York: Ginn and Company.

Calderon, C. I. 1956. "Fewest Words to Open the Widest Doors," Texas Outlook, 40:14-16.

Carroll, John B., P. Davies, and B. Richmond. 1971. Word Frequency Book. Boston: Houghton-Mifflin.

Card, William, and Virginia McDavid. 1965. "Frequencies of Structure Words in the Writing of Children and Adults," Elementary English, 42:878-882 and 894.

_____. 1966. "English Words of Very High Frequency," College English, 27:596-604.

Dale, Edgar. 1931. "Evaluating Thorndike's Word List," Educational Research Bulletin, 10:451-457.

_____. 1941. "A Comparison of Two Word Lists," Educational Research Bulletin, 10:484-489.

_____. 1956. "The Problem of Vocabulary in Reading," Educational Research Bulletin, 35:113-123.

_____. 1963. Bibliography of Vocabulary Studies. Columbus, Ohio: Bureau of Educational Research and Service, Ohio State University, Second Revised Edition.

Denslow, Orrience D. 1961. "Vocabulary and Sentence Study of Eight First Grade Science Books," Elementary English, 38:487-490.

Dolch, Edward William. 1927. "The Combined Word Study List," Journal of Educational Research, 17:11-19.

_____. 1928a. "Combined Word Studies," Journal of Educational Research, 17:11-19.

_____. 1928b. "Vocabulary Burden," Journal of Educational Research, 17:170-183.

_____. 1930. "Sampling of Reading Matter," Journal of Educational Research, 22:213-215.

_____. 1936a. "A Basic Sight Vocabulary," Elementary School Journal, 36:456-460.

_____. 1936b. "The First Step in Remedial Reading," Elementary School Journal, 37:268-272.

_____. 1939. A Manual for Remedial Reading. Champaign, Illinois: The Garrard Press.

_____. 1941. Teaching Primary Reading. Champaign, Illinois: The Garrard Press.

_____. 1942. Better Spelling. Champaign, Illinois: The Garrard Press.

_____. 1948. Problems in Reading. Champaign, Illinois: The Garrard Press.

_____. 1960. "Needed Vocabulary," Elementary English, 37:530-534.

Durr, William. 1973. "A Computer Study of High Frequency Words in Popular Trade Juveniles," The Reading Teacher, 17:1, October, 37-42.

Fennell, Ethel L. 1928. "Reocurring Words and Their Relation to Difficulties in Comprehension," Elementary School Journal, 29:42-53.

Fielstra, C. and F. D. Curtis. 1939. "Comparison of the Thorndike Word Lists," Journal of Educational Psychology, 30:445-452.

Fitzgerald, James. 1934. "The Vocabulary of Children's Letters, Written in Life Outside of School," Elementary School Journal, 34:358-370.

_____. 1938. "The Vocabulary and Spelling Errors of Third Grade Children's Life-Letters," Elementary School Journal, 38:518-527.

_____. 1958. "What Words Should Children Study in Spelling," Education, 79:224-228.

_____. 1963. "An Integrating Basic Communication Vocabulary," Elementary English, 40:283-289.

Francis, Nelson W. 1965. "A Standard Corpus of Edited Present-Day American English," College English, 26:267-273.

Fry, Edward. 1957. "Developing A Word List for Remedial Reading," Elementary English, 34:456-458.

_____. 1960. "Teaching A Basic Reading Vocabulary," Elementary English, 37:38-42.

Fullmer, D. W., and C. J. Kolson. 1961. "Beginning Reading Vocabularies," Journal of Educational Research, 54:270-272.

Gagg, J. C. 1959. "Beginning to Read: Choosing the Vocabulary," Times (London), Educational Supplement, 2298:1026.

Gates, Arthur I. 1926a. "Construction of a Reading Vocabulary for the Primary Grade," Teachers College Record, 27:625-642.

_____. 1926b. A Reading Vocabulary for the Primary Grades. New York: Teachers' College, Columbia University.

_____. 1935. Reading Vocabulary for the Primary Grades. New York: Teachers' College Press, Teachers College, Columbia University.

Gentry, Lillian. 1950. "A Study of the Vocabulary Load of Sixty-six Pre-primers," Journal of Educational Research, 43:525-532.

Gray, C. T. 1936. "A New Word List," Elementary School Journal, December, pp.556-557.

Gray, William. 1925, Twenty-fourth Yearbook of the National Society for the Study of Education. Part I. Bloomington, Illinois: Public School Publishing Company, pp. 193-198.

Hanna, P. R., Jean S. Hanna, R. G. Hodges, and E. H. Rudorf, Jr. 1966. Phoneme-Grapheme Correspondences as Clues to Spelling Improvement. Washington, D. C.: Office of Education, United States Department of Health, Education, and Welfare.

Harring, Sydney. 1931. "What Primer Shall I Use Next?" Elementary School Journal, 32:207-213.

Hayward, W. George, and Nancy M Ordway. 1937. "Vocabularies of Recently Published Pre-Primers," Elementary School Journal, 37:608-617. (As cited in Stone, 1949.)

Hildreth, Gertrude. 1948a. "Word Frequency As A Factor in Learning to Read and Spell," Journal of Educational Research, 41-467-471.

_____. 1948b. "A Comparison of the Dale, Dolch, and Rinsland Word List." Journal of Educational Psychology, 39:40-46.

_____. 1951. "Evaluation of Spelling Word Lists and Vocabulary Studies," Elementary School Journal, 51:254-265.

Hillerich, Robert L. 1974. "Word Lists-Getting It All Together," The Reading Teacher, Vol. 27, No. 4, January, pp. 353-360.

Hockett, John A., and Deta P. Neeley. 1936. "A Comparison of the Vocabularies of 33 Primers," Elementary School Journal, 37:190-201.

Hockett, J. A., and N. C. Neeley. 1937. "The Vocabularies of Twenty-Eight First Readers," Elementary School Journal, 37:344-352.

Hockett, J. A. 1938. "A Comparative Analysis of the Vocabularies of Twenty-Nine Second Readers," Journal of Educational Research, 31:665-671.

Horn, Ernest. 1925. "The Commonest Words in the Spoken Vocabulary of Children Up to and Including Six Years of Age," The Twenty-fourth Yearbook of the National Society for the Study of Education, Part I, Chapter VII, as cited in Buckingham and Dolch, 1936.

_____. 1926. A Basic Writing Vocabulary of 10,000 Words Mostly Used in Writing (Monograph). Iowa City, Iowa: University of Iowa, Monographs in Education.

Horn, Madeline D. 1926-27. "The Thousand and Three Words Most Frequently Used by Kindergarten Children," Childhood Education, 3:118-122.

_____. 1928. "A Study of the Vocabulary of Children Before entering the First Grade," 1201 16th Street, Washington, D. C.

Housh, E. T. 1918. "Analysis of the Vocabularies of Ten Second-Year Readers," The Seventeenth Yearbook, National Society for the Study of Education, Committee of Education, Part I, pp. 40-45.

Hughes, Marie M., and Vivian Cox. 1949. "The Language of First Grade Children, I and II," Elementary English, 26:373-380, 468-474, 495.

Irmina, Sister M., O.S.B. 1929. "An Evaluation of the Vocabulary Content of Twelve Series of Primary Readers," Educational Research Bulletin, Washington, D. C.: The Catholic University of America, Catholic Education Press, (5)7.

Johns, Jerry. 1971. "The Dolch Basic Sight Word List-Then and Now," Journal of Reading Behavior, Vol. 3, Fall, pp. 35-40.

_____. 1972. "The Dolch List Re-Examined-A Further Comment," Reading Teacher, Crossfire section, March, p. 533.

_____. 1975. "Dolch List of Common Nouns-A Comparison," Reading Teacher, Vol. 28, No. 6, March, pp. 538-540.

Johnson, Dale. 1971. "The Dolch List Re-Examined," The Reading Teacher, Feb., pp. 449-57.

Johnson, Glenn R. 1962. "Core Vocabulary Study in the Elementary Grades," Elementary English, 39:470-473.

Jones, W. F. 1915. A Concrete Investigation of the Material of English Spelling. South Dakota: University of South Dakota (as cited in Buckingham and Dolch, 1936).

Kirkpatrick, J. J. and E. E. Cureton. 1949. "Vocabulary Item Difficulties and Word Frequency," Journal of Applied Psychology, 33:347-351.

Kucera, Henry and W. Nelson Francis. 1967. Computational Analysis of Present-Day American English. Providence, Rhode Island: Brown University Press.

Kyte, George C. 1953. "A Core Vocabulary in the Language Arts," Phi Delta Kappan, 34:231-234.

Kyte, George C. and Virginia Keel. 1953. "Core Vocabulary of Spelling Words," Elementary School Journal, 54:29-34.

Loban, Walter. 1963. The Language of Elementary School Children. Champaign, Illinois: National Council of Teachers of English Research Report, No. 1, on Reading-Linguistics.

Lorge, Irving. 1944. "Predicting Readability," Teachers College Record, 45:404-419.

Lowe, A. J. and John Follman. 1974. "Comparison of the Dolch List with Other Lists," The Reading Teacher, Vol. 28, No. 1, October, pp. 40-44.

McKee, P. 1937. "Words Lists and Vocabulary Difficulties in Reading Matter," Elementary English Review, 14:241-245.

Paker, J. L. 1921. "The Vocabularies of Ten First Readers," The Twentieth Yearbook of the National Society for the Study of Education, Part II. Bloomington, Illinois: Public School Publishing Company, pp. 127-144.

Preiler, Arno G. 1967. "Some Problems Involved in Compiling Word Frequency Lists," The Modern Language Journal, 51:399-402.

Reeves, Olive R. 1958. "The Vocabulary of Seven Primary Reading Series," Elementary English, 35:237-239.

Rinsland, Henry D. 1945. A Basic Vocabulary of Elementary School Children. New York: Macmillan Company.

Seegers, J. C. 1940. "Vocabulary Problems in Elementary School: A Digest of Current Research," Elementary English Review, 17:28-43.

Selke, Erich, and G. A. Selke. 1922. "A Study of Vocabularies of Beginning Books in 12 Reading Methods," Elementary School Journal, 22:745-749.

Selke, E. 1929. "A Study of the Vocabulary of 10 Spellers," Elementary School Journal, 29:767-771.

_____. 1930. "A Comparative Study of the Vocabularies 12 Beginning Books in Reading," Journal of Educational Research, 22:369-374.

Stauffer, R. G. 1966. "Vocabulary Study Comparing Reading, Arithmetic, Health and Science Texts," The Reading Teacher, 20:141-147.

Stone, Clarence R. 1941. "Vocabulary of 20 Pre-primers," Elementary School Journal, 41:423-429.

_____. 1942. "A Vocabulary Study Based on 107 Primary Grade Books," Elementary School Journal, 42:452-455.

_____. 1950. Progress in Primary Reading. St Louis, Missouri: Webster Publishing Company.

Stone, David R., and Vilda Bartschi. 1963. "Basic Word List from Basal Readers," Elementary English, 40:420-427.

Strickland, Ruth G. 1962. "The Language of Elementary School Children: Its Relationship to the Language of Reading Textbooks and the Quality of Reading of Selected Children," Bulletin of the School of Education, Indiana University, Vol. 38, No. 4. Bloomington, Indiana: Bureau of Educational Studies and Testing, School of Education, Indiana University.

Taylor, Earl A. 1951. Basic Vocabulary for a New Education: Reading, Writing, Speaking Including the Components of a New Education. New York: Washington Square Reading Center.

Thorndike, Edward L. 1921a. The Teacher's Word Book of 10,000 Words, New York: Bureau of Publications, Teachers College, Columbia University.

_____. 1921b. "Word Knowledge in Elementay School," Teachers College Record, 22(4).

_____. 1931. The Teacher's Word Book of 20,000 Words. New York: Bureau of Publications, Teachers College, Columbia University.

_____. 1937. "The Vocabulary of Books for Children in Grades Three to Eight," Teachers College Record, 38: 196-205, 316-323, 416-429.

_____. 1940. "Value of Word-Counts: Critique of the Seventh Annual Research Bulletin of the National Conference on Research in English," Elementary English Review, 17:60-62,67.

_____. 1944. The Teacher's Word Book of 30,000 Words. New York: Bureau of Publications, Teachers College, Columbia University.

Tidyman, W. F. 1921. A Survey of the Writing Vocabularies of Public School Children in Connecticut. Washington, D. C.; United States Bureau of Education Teacher's Leaflet No. 15 (as cited in Buckingham and Dolch, 1936).

Wheeler, H. E. and Emma A. Howell. 1930. "A First-Grade Vocabulary Study," Elementay School Journal, 31:52-60.